THE UNIVERSITY *of* EDINBURGH

METHODS TO ASSESS CHANGES IN HUMAN BRAIN STRUCTURE ACROSS THE LIFECOURSE

David Alexander Dickie

PhD

The University of Edinburgh

2013

I declare that this work has not been submitted for any other degree or professional qualification and that it is my own, but where I acknowledge the contribution of others.

David Alexander Dickie
September, 2013

Acknowledgements

I wouldn't have been able to complete this work without my colleagues, friends, and family. Dr Dominic E. Job was my first supervisor. His great experience and insight improved this work endlessly. But more than that, he was my friend and got me through the darkest days of this process. Professor Joanna M. Wardlaw gave me the opportunity and funding to pursue this degree. Anything that I have achieved since then is mostly due to her direction (much needed prodding) and encouragement.

I received a considerable amount of support in the background of this process, the majority of which was beyond the call of duty. If I ever submitted a correct piece of paperwork it was due almost entirely to Moira "auntie Mo" Henderson and Dr Duncan J. Martin "OBE". They gave much more than that and I promise I will replace their Post-its one day.

My close lab mates (Andreas No Middle Name Glatz, Anna Barbara Jones, and Dr Jehill Parikh) helped me figure out a range of problems from computer codes to women codes. This process would have been a bigger struggle without them and am really glad we got to do it together. When I was away from the lab I was generally with the boys from home. They really did make me forget it all for a while and will never know how much that helped me through.

Lastly, I would be nothing without my family and owe anything good I've ever done to them.

Absent Friends.

For my Mum and Dad

Abstract

Human brain structure can be measured across the lifecourse ("*in vivo*") with magnetic resonance imaging (MRI). MRI data are often used to create "atlases" and statistical models of brain structure across the lifecourse. These methods may define how brain structure changes through life and support diagnoses of increasingly common, yet still fatal, age-related neurodegenerative diseases. As diseases such as Alzheimer's (AD) cast an ever growing shadow over our ageing population, it is vitally important to robustly define changes which are normal for age and those which are pathological. This work therefore assessed existing MR brain image data, atlases, and statistical models. These assessments led me to propose novel methods for accurately defining the distributions and boundaries of normal ageing and pathological brain structure.

A systematic review found that there were fewer than 100 appropriately tested normal subjects aged ≥60 years openly available worldwide. These subjects did not have the range of MRI sequences required to effectively characterise the features of brain ageing. The majority of brain image atlases identified in this review were found to contain data from few or no subjects aged ≥60 years and were in a limited range of MRI sequences. All of these atlases were created with parametric (mean-based) statistics that require the assumptions of equal variance and Gaussian distributions. When these assumptions are not met, mean-based atlases and models may not well represent the distributions and boundaries of brain structure.

I tested these assumptions and found that they were not met in whole brain, subregional, and voxel-based models of ~580 subjects from across the lifecourse (0-90 years). I then implemented novel whole brain, subregional, and voxel-based statistics, e.g. percentile rank atlases and nonparametric effect size estimates. The equivalent parametric statistics led to errors in classification and inflated effects by up to 45% in normal ageing–AD comparisons. I conclude that more MR brain image data, age appropriate atlases, and nonparametric statistical models are needed to define the true limits of normal brain structure. Accurate definition of these limits will ultimately improve diagnoses, treatment, and outcome of neurodegenerative disease.

Contents

Glossary of terms

2D	Two dimensional
3D	Three dimensional
4D	Four dimensional
AD	Alzheimer's disease
ADNI	Alzheimer's Disease Neuroimaging Initiative
ANTS	Advanced Normalization Tools
BET	Brain Extraction Tool
BRAINS	Brain Images of Normal Subjects
CSF	Cerebrospinal fluid
CDR	Clinical dementia rating
CT	Computed tomography
CV	Coefficient of variance
DICOM	Digital Imaging and Communications in Medicine
DTI	Diffusion tensor imaging
FAST	FMRIB's Automated Segmentation Tool
FMRIB	Oxford Centre for Functional Magnetic Resonance Imaging of the Brain
FID	Free induction decay
FLAIR	Fluid Inversion Attenuated Recovery
FSL	FMRIB's Software Library
GLM	General linear model
GM	Grey matter
ICBM	International Consortium for Brain Mapping
IQR	Interquartile range
LONI	Laboratory of Neuro Imaging
MATLAB	Matrix Laboratory
MMSE	Mini–mental state examination
MNI	Montreal Neurological Institute
MP-RAGE	Magnetization prepared rapid gradient–echo
MRI	Magnetic resonance imaging
MTL	Medial temporal lobe
NART	National Adult Reading Test
OASIS	Open Access Series of Imaging Studies
ROI	Region of interest

SD	Standard deviation
SNR	Signal-to-noise ratio
SPM	Statistical Parametric Mapping
T	Tesla
T1	Longitudinal relaxation time
T2	Transverse relaxation time
TIV	Total intracranial volume
WAIS	Wechsler Adult Intelligence Scale
WM	White matter
WMH	White matter hypointensities/ hyperintensities
WMS	Wechsler Memory Scale

Symbols

β	Beta coefficient (from regression analysis)
d	"Cohen's d" (measure of effect size)
δ	Effect size delta
μ	Mean
Δ_μ	Mean change
/	Or
p	Percentile rank
σ	Standard deviation

List of tables

List of figures

1. Introduction and motivation

With 100 billion neuronal cells and up to 1000 trillion synaptic connections between cells, the human brain is the most complex structure in the known Universe (Koch, 2004, Pakkenberg and Gundersen, 1997, Williams and Herrup, 1988). This number of component parts (10^{26}) is approximately the number of stars in the entire observable Universe and 10 times the number of grains of sand on planet Earth (Derbyshire, 2003; van Dokkum and Conroy, 2010; European Space Agency, 2013). It is this great complexity that most differentiates humans from other animals and is the source of our unique intellectual ability (Jones, 2012). This intellectual ability has driven over 5000 years of neuroscientific study, allowing us to somewhat understand and react to changes and diseases that occur in the brain during the human lifecourse (Adelman and Smith, 1987; Job et al., 2002; Fox and Schott, 2004; Wardlaw et al., 2009).

For example, as our population ages the growing incidence of neurodegenerative diseases such as Alzheimer's (AD) may be stemmed if subtle pathological changes in the brain are noticed early enough (Brody, 1985; Selkoe, 2001; Selkoe, 2013). However, not all changes in brain structure indicate neurodegenerative disease. We still do not fully understand the pathological changes that lead to these increasingly common, yet still fatal, diseases (Shenton et al., 2001; Farrell et al., 2009; Ferguson et al., 2010; Davenport, 2013). If pathological changes are not accurately identified in the incipient stages then treatments can do little more than preserve an already diminished state of being (Selkoe, 2001; Fox and Schott, 2004; Selkoe, 2013). Although diseases like AD are currently incurable, if these changes are identified in the incipient stages then their symptoms may be slowed or even stopped (Selkoe, 2013).

The present work therefore aims to contribute to better understanding of changes that occur in the human brain across the lifecourse. This is in turn may circumvent the shadow of neurological disease that looms over our ageing population (Brody, 1985; Wardlaw et al., 2009; Selkoe, 2013). As was noted in 3000 B.C. (Adelman and Smith, 1987), the first step to providing this understanding is to accurately measure the structure of the brain. The first sections of this thesis describe key methods for providing these measurements.

1.1 Methods for assessing brain structure

The gyral and sulcal pattern of the human brain was first described in 3000 B.C. (Adelman and Smith, 1987). Although study of the structure of the brain continued for more than 4500 years, it was not until 1664 when Thomas Willis published *Cerebri Anatome* ("Anatomy of the Brain") that robust methods for measuring brain structure were developed (O'Connor, 2003). Willis directed novel autopsies of the brain in which it was first removed from the skull, in contrast to the traditional *in situ* dissections of the time, and then sliced from the base upwards. These slices were then viewed with a microscope and drawn by Christopher Wren (O'Connor, 2003). This method, inherent in modern "*in vivo*" imaging (section 1.1.1), led to the discovery of the "circle of Willis", the joining area of several arteries that collectively supply oxygenated blood to over 80% of the brain (Jasmin and Zieve, 2012).

Following Willis and Wren, the first major study of the brains' gyral and sulcal pattern ("cranio–cerebral topography") was undertaken in 1876 by Paul Broca. By drilling metal pins between them, Broca identified spatial relationships between sulci and cranial (skull) sutures (Broca and Brabrook, 1881; Anderson and Makins, 1889). These early brain "atlases" were later advanced by Korbinian Brodmann in 1909. Brodmann used microscopy to define 43 distinct brain regions (now known as "Brodmann's areas") based on their cellular organisation (Brodmann, 1994; Zilles and Amunts, 2010).

The "circle of Willis", "Broca's area", and all 43 of "Brodmann's areas" form part of the lexicon of modern neuroscience. The methods used to derive them (e.g., dissection, drilling, and microscopy) are, however, not entirely suitable for modern day *in vivo* ("in the living") studies. To better understand how the brain changes during life and the effects these changes have on function and health, less archaic methods of studying the brain were required. It was the accidental discovery of the ionizing radiation (X)–ray in 1895 by Wilhelm Conrad Röntgen that led to the development of these methods (Bull, 1961).

1.1.1 *In Vivo* brain imaging

Following the discovery of the X–ray in 1895 by Wilhelm Conrad Röntgen, the first radiological diagnosis of a pituitary tumour was made in 1897 (Bull, 1970). For the next 70 years X–ray brain imaging became very useful in diagnosing most types of

hematomas (bleeding) and hemangiomas (tumours). However, the contrast in "whole head" X–ray imaging meant that pathology could not always be differentiated from normal tissue and the macro brain structures identified in Wren and Brodmann's post mortem drawings were not yet distinguishable *in vivo*.

It was the development of "planar" X–ray imaging, or X–ray computed tomography (CT), in 1972 by Sir Godfrey N. Hounsfield that provided the means for reliable differentiation between normal and pathological brain tissue (Hounsfield, 1973; Isherwood, 2005). Just as Willis was able to accrue more detail about the brain by slicing it in sections with a blade, CT directed X–rays in cross-sectional "slices" of the brain to provide more detail (than the previous whole head X–rays). The varying rates of absorption ("Hounsfield units") of X-rays in different structures (e.g. tissue, fat, fluid …) produced high contrast, cross-sectional images. These two dimensional (2D) images were then reconstructed to create equally high contrast three dimensional (3D) images of the whole head (Hounsfield, 1973).

Despite this great advance, the contrast between tissues in brain CT images was sometimes limited. Moreover, by exposing patients to ionizing radiation, the 72 million CT scans performed in the United States in 2007 may contribute to up to 45,000 future cancer cases (Berrington de Gonzalez et al., 2009). Further developments were then clearly required to safely capture the detail of Wren, Broca, and Brodmann *in vivo*. This method arrived through observation of nuclear magnetic resonance (NMR) in 1946 (Bloch, 1946). NMR was the basis for what became magnetic resonance imaging (MRI).

1.1.2 Magnetic resonance brain imaging

In 1946, Felix Bloch and Edward Purcell used magnetic fields and radiofrequency to observe the composition of molecular structures (Bloch, 1946; Hofstadter, 1994). But it was not until 1973 that Paul Lauterbur used NMR to image chemical compounds by exploiting the tiny electrical current given off water when it is placed within a magnet and exposed to a radio wave (Lauterbur, 1973). Lauterbur's "zeugmatogram" (internal image) of a glass capillary of water was followed by a zeugmatogram of a human finger in 1976 by Peter Mansfield (Mansfield and Maudsley, 1977). Although the term is no longer widely used, the principles of zeugmatography form the basis for all modern brain MRI (Figure 1).

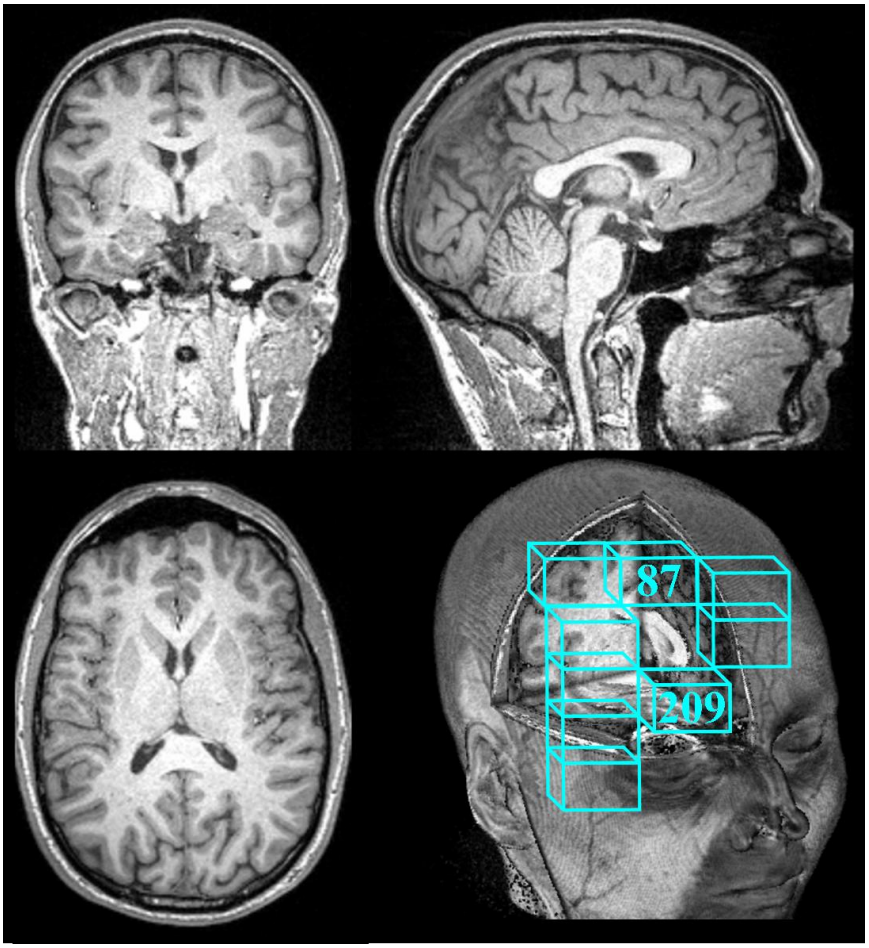


Figure 1. A modern longitudinal relaxation time (T1)–weighted MR brain image

This shows coronal (top left), sagittal (top right), axial (bottom left), and 3D rendering (bottom right) views. The lattice structure in the 3D rendering is an exaggerated illustration of the numerical voxel matrix that the image consists of (the actual voxel size is $1mm^3$). Although there is no standard meaning on MRI intensities, the values 87 and 209 provide an example of the ratio between cerebrospinal fluid and grey matter values.

The potential of zeugmatography to provide detailed images of the brain *in vivo* was described in 1978 by Mansfield (Mansfield and Pykett, 1978) and the first MR brain images, derived in a magnet of 0.1 tesla (T; which is still some 15,000 times the strength of the Earth's magnetic field), were published in 1980 (Holland et al., 1980; Hoeffner et al., 2012). These images are essentially matrices of numbers (intensities) and individual elements within these matrices are known as "voxels", i.e. 3D pixels (Figure 1). Unlike CT images (that are based on the standard Hounsfield scale), there is not a standard MR intensity scale.

Despite the lack of a universal intensity scale, many research centres started to invest in MRI scanners and produce high resolution images of the brain. Although

many features, such as such as the ventricular system, Sylvian fissure, interhemispheric fissure and tissue boundaries, were visible in CT, the clarity of detail previously drawn *ex vivo* by Wren and Brodmann, was now provided *in vivo* with MRI (Figure 1) (Hawkes et al., 1980). During the thirty years following 1980, MRI was developed to produce high resolution images of the brain that clearly defined subregional structures as small as the hippocampus (~7ml), amygdala (~5ml), and third ventricle (~1ml) (Filipek et al., 1994). The human brain has now been imaged with a magnet of up to 9.4 T and in–plane resolution (voxel size) of *in vivo* images may now be as low $0.5mm^2$ (Vaughan et al., 2006). With these parameters the fibres of the white matter and subfields of subregional structures may now be clearly defined, e.g. hippocampal subfields, that could only previously be viewed in histology (Mueller et al., 2007).

This range of detail is provided by the array of MRI "sequences" (timings and directions of radio frequency pulses and signal capture) that have been developed. A non-exhaustive, but commonly used list is given in Table 1.

Table 1. Structural MRI sequences

MR sequence name	Example brain features
Diffusion tensor imaging (DTI)	White matter fibre tracks
Diffusion weighted imaging (DWI)	Ischaemic parenchyma
Fluid attenuated inversion recovery (FLAIR)	Periventricular white matter and hyperintensities
MR angiogram	Blood vessels, e.g. "circle of Willis"
Longitudinal relaxation time (T1)	Cortical and subcortical volumes, e.g. post central gyrus, hippocampus
T1 gradient echo (GRE)	Cortical and subcortical volumes with reduced scan time and noise
Transverse relaxation time (T2)	White matter hyperintensities
T2* GRE	Brain boundary (mask), haemorrhage

Note: This table was derived from Farrall (2006), McRobbie et al. (2006), and Pooley (2005).

These technological developments and subsequent potential to better understand how brain structure changes during life (discussed further in section 1.2) led to a dramatic increase in brain MRI studies with more than 144,000 research subjects scanned between the early 1990s and October 2011 (Poline et al., 2012). In the earliest of these years, and even more recently, most of these data were analysed and then archived

after their primary purpose was served. This may seem like a reasonable thing to do and not at all wasteful, but this is unfortunately not the case. For example, if no further uses of petroleum were sought beyond its' primary product (motor gasoline), it would cost the oil industry US$1,150,000,000,000 (US$1.15 trillion) per year (Ross, 2012).

Although they may not directly provide financial profit, brain MRI data are very valuable and have the potential to answer unforeseen questions beyond those originally posed (Koslow, 2000; Toga, 2002). This has been recognised by research funding organisations such as the Medical Research Council (MRC) who now require scientific data to be openly shared where possible after their initial purpose has been served (Arzberger et al., 2004). Therefore, to exploit their full value and potential to provide a deeper understanding of the brain, databanks (rather than inaccessible archives) of MR brain images were established.

1.1.3 Magnetic resonance brain image databanks

MR brain image databanks are structured and accessible collections of the primary image data as well as the associated metadata, e.g. patient characteristics and sequence parameters. The metadata are essentially held in spreadsheets (tables) and each subject record (spreadsheet row) is linked to their corresponding images that are independently held on a computer drive or server. This linkage allows users to search for images by the characteristics (spreadsheet columns) that makeup each subject record, e.g. if such metadata are provided by the databank, a list of subjects and their corresponding images may be generated by entering the search: MRI sequence=T1; age=60 years; gender=female; systolic blood pressure=150 (Marcus et al., 2007c).

When MR brain image data are made accessible to others in this manner, it allows them to validate results obtained from their own data, perform pilot studies, develop standardised procedures, produce realistic teaching material, and potentially provide a more integrated view of brain structure, function, and health (Koslow, 2000; Toga, 2002). More specifically, if these databanks are made available to others they will provide a source of images to "atlas" the brain which, as demonstrated by Willis, Broca, and Brodmann, has become a reliable method for better understanding it.

1.1.4 Magnetic resonance brain image atlases

Brain atlases, through the work of Willis, Broca, Brodmann and more recently, Duvernoy, Ono, and Talairach, have been in development for hundreds of years (Broca and Brabrook, 1881; Talairach and Tournoux, 1988; Ono et al., 1990; Brodmann, 1994; Duvernoy, 1999; O'Connor, 2003). The earliest atlases were based on drawings and later film photographs of one or few brain specimens. They were created to map the structure of the brain, correlate these structures with specific functions, and guide cranial surgeries (Broca and Brabrook, 1881; Anderson and Makins, 1889; Talairach and Tournoux, 1988). But, as their creators themselves noted (Talairach and Tournoux, 1988), these atlases were not able to adequately describe the high variance in brain structure between individuals, even those of the same gender and age (Farrell et al., 2009). The advent of MR imaging and later brain MRI databanks allowed the relatively quick acquisition and storage of a mass of brain images (relative to the painstaking patience and physical storage required for drawings and film photographs). Furthermore, MR brain images could be much more easily combined, i.e. the mean of all subject intensities in each voxel (Figure 2), to create composite images that better reflected the variance in brain structure (Evans et al., 1993; Evans et al., 2012).

The first MR brain image atlases[1] were derived from the Talairach and Tournoux specimen atlas and this is still the basis for many modern atlases (Figure 2). The Talairach and Tournoux atlas itself was derived from the brain in a 60 year old female cadaver (Talairach and Tournoux, 1988). One hemisphere was sectioned in the sagittal plane and the other in the coronal plane and from these sections the axial images were derived, ignoring hemispheric asymmetry (Talairach and Tournoux, 1988; Evans et al., 2012). These images were digitised to generate the electronic, "Talairach space".

In an attempt to further knowledge derived from atlases based on one or few individuals, the Montreal Neurological Institute (MNI) calculated the mean intensity

[1] Note that the term "brain image atlas" is sometimes distinguished from the term "brain image template" in the literature. That is, where a "template" may consist of brain structure derived from one or many individuals in a standard framework; an atlas, while still consisting of brain structure derived from one or many individuals in a standard framework, also consists of labels that have been added to specific regions to identify them as, for example, hippocampal grey matter or lateral ventricle cerebrospinal fluid. However, others do not make this distinction between the terms and use them interchangeably. In this work we too use these terms interchangeably, generally preferring the term "atlas".

image from 305 MRI subjects that had been “registered”, i.e. overlaid, into Talairach space (Evans et al., 1993). This “MNI305” atlas was the basis for the “MNI152” atlas which was the mean of 152 young adult MRI subjects in MNI305 space (Figure 2).

These modern atlases are mainly used in functional brain imaging studies where the mean blood flow of several subjects in the atlas space is correlated with specific functions. This, for example, is done to validate the structure–function relationships that were first identified with saws and drills by Broca and Brodmann (Hinke et al., 1993; Buccino et al., 2001). In order to achieve these results in a functional brain imaging study, each subject must be overlaid (“registered”) to the space defined by the atlas. That is, the angle, position, size, and shape of each subject brain are deformed to those of the atlas (Figure 2). This is also done in structural brain image studies but this degree of registration may not be appropriate for studies that are concerned with defining the variation in brain structure between individuals. As shown in Figure 2, each subject’s brain structure is warped to the structure of the atlas, removing almost all of the structural variance between subjects.

Some of the variance between subjects, e.g. head size, is not strongly associated with age (after childhood) or dementia (although it is moderately associated with intelligence; Courchesne et al., 2000; Good et al., 2001; Edland et al., 2002; MacLullich et al., 2002). For this reason, relative measurements of brain structure, e.g. head size normalised tissue volumes, are calculated when attempting to explain variation between age or disease groups (Fotenos et al., 2005). Relative measurements are calculated by matching only the angle, position, and size of subject brains, i.e. maintaining inter–subject shape differences (in voxel–based studies), or by calculating subject volumes of interest, e.g. hippocampus volume, as a proportion of their total intracranial volume (Walhovd et al., 2005a). Following these image processing steps, the associations between brain structure and variables of interest, e.g. age, may be determined with statistical models.

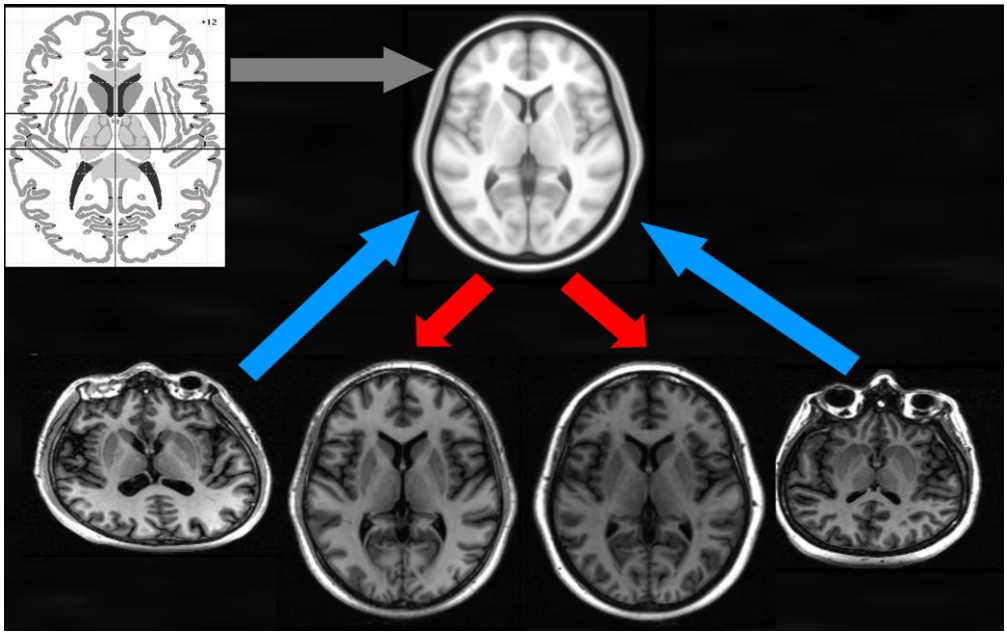


Figure 2. Brain atlases and subject registration
The Talairach brain specimen atlas (top left*), Montreal Neurological Institute (MNI) 152 brain image atlas (top right), an original individual subject (bottom left), and the same subject registered to the MNI152 atlas. The registration has altered the morphology of the original subject to that of the template, e.g. the sulcal spaces and lateral ventricles have been reduced. *Copyright © 2003-2013 Research Imaging Institute, University of Texas Health Science Center at San Antonio.

1.1.5 Statistical models of magnetic resonance brain images

Statistical modelling is used to determine relationships between dependent variables, e.g. brain tissue volume, and independent variables, e.g. age, cognitive state. There are a number of approaches to statistical modelling, e.g. graphical modelling, parametric modelling, nonparametric modelling, hierarchical Bayesian modelling (Freedman, 2010). The list is extensive but the most common approach in structural brain image analysis is parametric modelling.

Parametric modelling itself has a number of subcategories, e.g. tests of group means (*z*– and *t*–testing) versus tests of variances (Hogg and Tanis, 2010). The most commonly used parametric approach is the general linear model (GLM; which often encompasses the t– and F–tests of means, e.g. Good et al., 2001). The GLM (the basis for linear regression and analysis of variance; ANOVA) has been used to infer general relationships between brain structure and independent variables, e.g. age/ gender, for over twenty years and is still commonly used at present (Gur et al., 1991; Good et al., 2001; Ziegler et al., 2011; Li et al., 2012; Long et al., 2012; Peelle et al., 2012).

The GLM is summarised by equation 1,

$$m = \frac{rSD_y}{SD_x} \tag{1.1}$$

$$\hat{y} = mx + c \tag{1.2}$$

where m is the slope of the regression line (also known as the beta (β) coefficient, e.g. mean brain tissue loss per year), r is the correlation between the variables x (e.g. age) and y (e.g. brain size), SD_y is the standard deviation of y, SD_x is the standard deviation of x, $\hat{y}$ is the GLM predicted value of brain size, x is the actual value of age, and c is the intercept of the regression line (the predicted value for y when x is 0) (Freedman et al., 2007).

The main outputs from GLM studies are the parameters derived from equation 1 (m/β and c), the "test statistic" from the hypothesis test that the equation was produced by chance (i.e. rather than the alternative hypothesis that there is actually an effect of x on y), and the resultant probability that the equation was produced by chance (the so called "*P*–value"). The m/β coefficient of the GLM is essentially the estimated mean difference in a dependent variable across the independent variable, e.g. the mean difference in brain tissue volume across age.

It has been recommended by several prominent statisticians that the GLM should only be used when the data meet a number of assumptions about their generation and distribution (Meehl, 1978; Cohen, 1994; Freedman et al., 2007; Freedman, 2010). Firstly, the subjects should be generated from a statistically random sample of the larger population of interest. A statistically random sample is equivalent to "blindly drawing tickets from a box" where the entire contents of the box, that each have an equal probability of selection, are the population and the tickets drawn are the sample (Freedman et al., 2007). It is very difficult to generate a statistically random sample of human subjects; the use of advertisements in hospitals to recruit subjects does not replicate a statistically random process. For example, people who do not see the advert have a zero probability for selection, there are non-random reasons why the eventual subjects were in the hospital to see the advertisement, and these subjects also had the choice whether or not to respond to the advertisement (Meehl, 1978). To reliably validate results when data are not random (i.e., recruited with selection bias), statistical models should be repeated in other samples and the primary measures of

differences, e.g. *m/β* coefficients (not *P*–values), obtained in the separate studies should be compared (Meehl, 1978; Cohen, 1994; Freedman, 2010).

Further to being derived randomly, data are also assumed to be distributed "equally Gaussian" (statistically normal) between each value of the independent variable. That is, the values of the dependent variable, e.g. *y*, within each value of the independent variable, e.g. *x*, should be distributed approximately equally to the Gaussian distribution and the variance between values of *x* should be equal (*homoscedastic* – literally translated as "of the same spread") (Freedman et al., 2007). An annotated illustration of the Gaussian distribution is shown in Figure 3 and examples of *homoscedastic* and *heteroscedastic* (unequal variance) data are shown in Figure 4. "Random effects" (controlling variables with various levels that are not all included in the model, e.g., only "low" or "high" blood pressure) may adjust for lack of homoscedasticity. That is, brain tissue volume distributions of subjects with only high or low blood pressure may be less skewed (den Heijer et al., 2005).

Percentile ranks are levels that represent percentages of subjects within a distribution, e.g. the bottom 5% of subjects in a distribution have a value equal to or less than the 5^{th} percentile rank value (Freedman et al., 2007). Figure 3 shows that in Gaussian distributed data, the mean minus 2 standard deviations (SD) is approximately equal to the 2.5^{th} percentile rank; the mean minus 1 SD is approximately equal to the 16^{th} percentile rank; the mean is equal to the 50^{th} percentile rank (median); the mean plus 1 SD is approximately equal to the 84^{th} percentile rank; and the mean plus 2 SD is approximately equal to the 97.5^{th} percentile rank (Freedman et al., 2007). The exact values for the 2.5^{th} and 97.5^{th} percentiles are $\mu \pm 1.96\ \sigma$ (from the Gaussian distribution lookup table).

When data are distributed as in the top panel of Figure 4 the *m/β* coefficient is generalisable across the distribution of data, i.e. subjects that are far from the mean differ between groups as much as subjects that are close to the mean. Brain image data, however, are often not distributed in this manner (Rorden et al., 2007). In particular, there are often skewed distributions of imaging data (bottom panel in Figure 4) in diseased and disordered populations such as AD and schizophrenia (Shenton et al., 2001; Barkhof et al., 2011). This means that, while it has successfully defined mean changes in brain structure, the GLM and other parametric methods may not be the most appropriate method for defining the other levels and limits of brain

structure (Freedman et al., 2007). True definitions of the limits of brain structure are very important to obtain since people at these limits may be at increased risk of developing overt neurodegenerative disease. Further, subjects who later developed neurodegenerative disease could be retrospectively removed from models to determine whether they were responsible for skewed distributions and limits.

As data accumulate, these levels and limits may be better defined with nonparametric statistics. However, the mean estimates, i.e. *m/β* coefficients, derived from the GLM still provide general definitions of how the brain changes across the lifecourse and these are described in the next section.

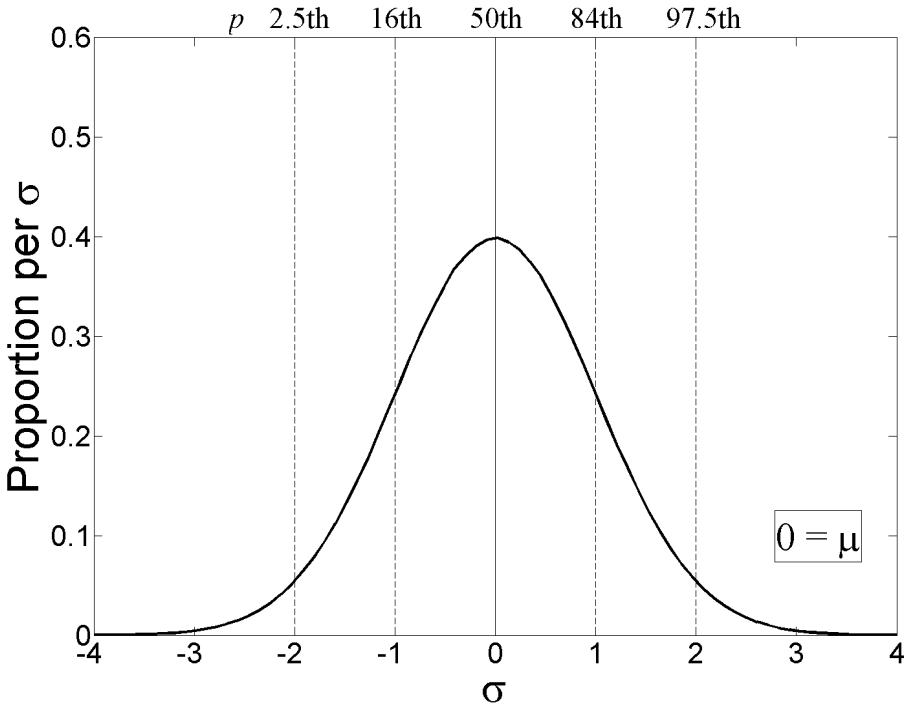


Figure 3. The Gaussian distribution
σ=standard deviation; μ=mean; *p*=percentile rank.

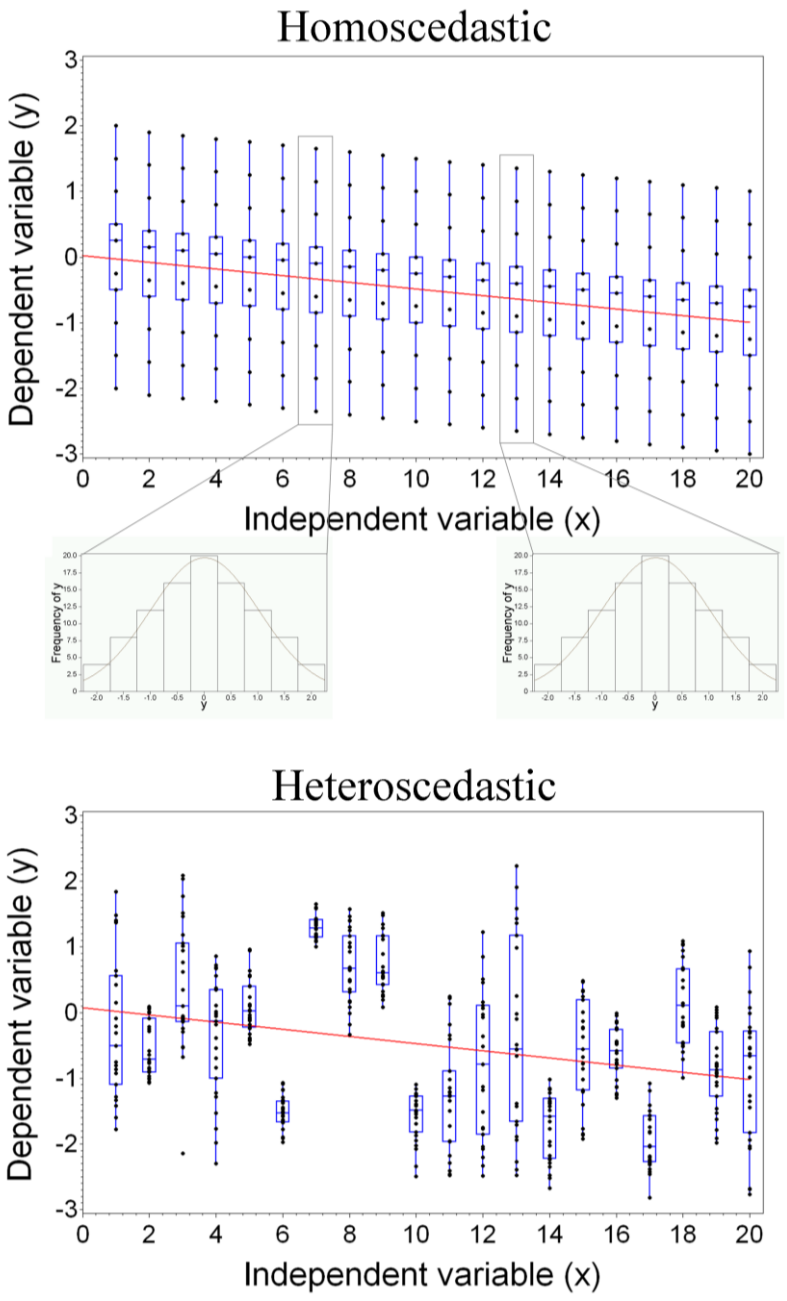


Figure 4. Example of equally Gaussian and unequally distributed (*heteroscedastic*) data with similar general linear model results (both $m/\beta \cong 0.05$)

In the equally Gaussian data (top panel) the values of y in each value of x (individual subjects are shown by the dots) are distributed approximately Gaussian and the variances between values of x are equal (*homoscedastic*). Equality of variance between values of x is shown by the equally sized box plots. The superimposed outline of the Gaussian distribution on the data histograms in middle panel show the conformance of the data to the Gaussian distribution. In contrast, the varying sizes of the box plots in the bottom panel show that these data are *heteroscedastic* (have unequal variance).

1.2 General patterns of brain structure development, ageing, and disease

Through post–mortem brain measurements, *in vivo* imaging techniques, databanks, atlases, and statistical models of these data we have developed a reasonable understanding of how the brain, in general, changes across the lifecourse and disease.

The size and shape of the brain changes from gestation until death, with the greatest changes occurring in early and late life (Scahill et al., 2003; Sowell et al., 2004; Fotenos et al., 2005; Prayer et al., 2006). These longitudinal changes are often inferred from cross–sectional differences between ages and, despite potentially confounding factors, e.g. technological developments and differences in nutrition across generations, these two types of study generally agree (Fotenos et al., 2005). From cross–sectional studies of whole brain volumes, brain tissue volume is said to increase rapidly after birth until adolescence and then decline steadily until old age, when it then declines more rapidly (Courchesne et al., 2000). The greatest tissue volume change is said to be in the grey matter whereas white matter volume is generally more constant across the lifecourse (Good et al., 2001). In region of interest (ROI) studies, age–related changes in volume and voxel tissue proportions are said to be more pronounced in the lower part of the brain, e.g. medial temporal lobe, than in the upper part of the brain, e.g. parietal lobe (Allen et al., 2005).

These age–related changes were calculated with subjects from across the lifecourse however, despite an ageing population (Brody, 1985), there was often a limited number (<100) of older subjects (>60 years) in these studies (Courchesne et al., 2000; Good et al., 2001; Sowell et al., 2003; Walhovd et al., 2005a; Giorgio et al., 2010; Li et al., 2012; Long et al., 2012). The variance in brain structure between older subjects has often been shown to be greater than in younger subjects (Manolio et al., 1994; Ge et al., 2002; Resnick et al., 2003; Kruggel, 2006; Farrell et al., 2009). This means that these general changes may not be applicable to the limits of normal ageing brain structure. In other words, those at the clinical limits may change at a different rate to those close to the mean (Elveback et al., 1970; Freedman et al., 2007), and therefore may be more susceptible to age–related cognitive decline and neurodegenerative disease.

Diseases such as Alzheimer's are thought to have a devastating but focused effect on brain structure. Through histological and imaging studies, a general pattern of brain structure changes has been shown to differentiate AD from normal ageing

(Braak and Braak, 1997; Jack Jr et al., 1997; Thompson et al., 2003; Fox and Schott, 2004). At disease onset there is thought to be accelerated loss of neurons (and subsequently tissue) in the hippocampus which is then focused throughout the medial temporal lobe (MTL). However, there is a wide range of hippocampus tissue loss in normal ageing and, although certainly a reasonably strong indicator, hippocampal atrophy alone may not be pathological (Ferguson et al., 2010). There is also a general trend for lateral ventricle expansion in AD (Thompson et al., 2003; Thompson et al., 2004). Relative to the strong (mean-based) association between AD and hippocampal atrophy/ ventricular expansion, frontal lobe atrophy is not thought to be greater in AD until the late stages of disease (Geroldi et al., 1999; Thompson et al., 2003). Further, many of the superior structures, e.g. sensory motor cortex, are thought to be spared in AD (Braak and Braak, 1997; Thompson et al., 2003).

This pattern of brain structure changes provides a general description of pathology progression however, it is not always replicated clinically where a range of atrophy is found across normal ageing and disease (Farrell et al., 2009), e.g. frontal atrophy is often seen in AD patients. The lack of research-based evidence for these common clinical findings suggests that alternative statistical models may be required to support diagnoses of increasingly common diseases such as Alzheimer's.

1.3 The growing shadow of neurodegenerative disease

Our population is ageing and this is causing a rapid increase in the rates of cognitive decline and subsequent neurodegenerative disease such as AD and other dementias (Brody, 1985; Wardlaw et al., 2009; Luengo-Fernandez et al., 2010; Selkoe, 2013). AD is characterised clinically by progressive cognitive and physical decline which eventually may require fulltime and prolonged nursing care (Blennow et al., 2006). Although only mild at first, by the time of even the earliest clinical symptoms there has already been silent but irreversible damage done to the brain (Selkoe, 2001; Blennow et al., 2006; Tondelli et al., 2012). Once initiated this brain damage causes a devastating cascade of tissue loss and degrading physical and cognitive abilities.

From the first diagnosis to the final mental and physical incapacity, these diseases are emotionally and financially devastating to patients, families, and carers. People cannot cope with this on their own and require help from their governments

but the growing incidence of cases means that, if methods for earlier diagnoses are not found, the United Kingdom government (and those across the world) will be crippled by nursing care costs running into hundreds of billions (Luengo-Fernandez et al., 2010). The estimated worldwide cost of dementia in 2010 was US$604 billion (Alzheimer's Disease International, 2013). This is almost twice the amount of market capital of the world's largest public limited company (Telegraph Media Group Limited, 2013). These overwhelming human and economic costs have made research into brain ageing, and methods for better diagnoses of age–related neurological disorders, a UK government and worldwide priority (Luengo-Fernandez et al., 2010; Collins and Prabhakar, 2013; Schwartz, 2013).

One way to meet this priority is to define changes in the brain that are normal for age and those that are indicative of cognitive decline and neurological disease. It is therefore the aim of this thesis to contribute to a better understanding of normal and pathological brain structure changes.

1.4 Aims of thesis

To better understand normal and pathological brain structure changes it is apparent that large volumes of data, representative and robust atlases, and generalisable statistical models are required. I therefore investigated these existing methods and set the following hypotheses and sub aims:

A1. Determine whether there are sufficient brain MRI data available to understand ageing brain structure adequately

H1. Parametric atlases adequately describe the distribution of brain structure across the lifecourse:

A2. Provide representative and robust atlases of the distribution of brain structure across the lifecourse

H2. Parametric statistical models adequately describe the distributions and boundaries of brain structure in normal ageing and disease:

A3. Provide generalisable statistical models of the distributions and boundaries of brain structure for assessments of groups and individuals

My contribution to meeting these aims is summarised in section 10.1, "Contribution".

1.4.1 Determine whether there are sufficient brain MRI data available to understand ageing brain structure adequately

"*Data! data! data!" he cried impatiently. "I can't make bricks without clay*."

–Sherlock Homes (Doyle, 1892)

As noted in section 1.1.3, the need for data to fully understand brain structure was well known (Koslow, 2000; Toga, 2002). However, it was not known whether sufficient normal ageing brain image data were currently available. Therefore, the first aim of this thesis was to determine whether or not there were sufficient data available to define changes in the brain that are normal for age and those that are indicative of cognitive decline and neurological diseases such as AD.

1.4.2 Provide representative and robust atlases of the distribution of brain structure across the lifecourse

"If our brains were so simple that I could understand them, I would be so simple that I could not."

–Anonymous (Lezak et al., 2004)

The complexity of the human brain means that atlases based on few individuals or the averages of even large amounts of MRI data may not adequately describe its structure (Rorden et al., 2007; Evans et al., 2012). Therefore, my second aim was to test

existing methods for creating atlases and then to create an atlas that was representative and robust to the complex and variable distributions of brain structure.

1.4.3 Provide generalisable statistical models of the distributions and boundaries of brain structure for assessments of groups and individuals

> "*It is difficult to understand why statisticians commonly limit their enquiries to Averages, and do not revel in more comprehensive views. Their souls seem as dull to the charm of variety as that of the native of one of our flat English counties, whose retrospect of Switzerland was that, if its mountains could be thrown into its lakes, two nuisances would be got rid of at once.*"
>
> –Sir Francis Galton, 1889 (Freedman et al., 2007)

The GLM and other parametric methods have successfully defined general (mean) brain structure changes across the lifecourse. However, if the assumptions of parametric statistical models are not met, their results will likely not be generalisable to subjects that are far from the mean value, i.e. at clinical limits (Elveback et al., 1970; Freedman et al., 2007). And, as discussed above, the assumptions of parametric statistical models may not always be met in MR brain structure data. Therefore, my final aim was to develop statistical models that were generalisable to subjects across the distributions of normal and pathological MR brain structure.

1.5 Summary of chapters

In this thesis I will first summarise each of the following nine chapters.

1.5.1 Chapter 2

Chapter 2 is a systematic review of structural brain MRI databanks with normal aged (≥60 years) subjects. Despite initially appearing to be several hundred, there were only 98 (appropriately determined) normal aged subjects openly accessible worldwide and these only had T1–weighted images.

1.5.2 Chapter 3

Chapter 3 is a formal review of structural brain MRI atlases. Most atlases were originally derived from the Talairach and Tournoux brain specimen atlas and consisted of mainly young adult subjects. All but one atlas was based on the mean of subject values, and there was no electronic atlas of the normal range of brain structure for any age. There were three atlases that spanned the adult lifecourse, e.g. ~18–90 years, of which the largest had 40 subjects. There were no electronic atlases for focused, e.g. <5 year interval, age ranges in older subjects.

1.5.3 Chapter 4

Chapter 4 is a formal review of statistical models of normal brain volumes across the lifecourse. I found one previous statistical model of brain structure that attempted to define differences in the limits of brain structure between ages. These limits were based on the equally Gaussian assumption of mean-based (parametric) models. All other studies provided mean only estimates of brain structure changes through life.

1.5.4 Chapter 5

Chapter 5 describes the image processing methods I chose to employ and the assessments I conducted to inform these choices. I developed and validated image processing pipelines for infants, young to middle (25–60 year) adults, and aged (60–90 year) adults. This included implementation and validation of a novel registration method, "nonlinear surface" registration to fully account for head size differences and maintain within brain variance of interest; and MR intensity normalisation for combining data from multiple scanners and acquisition parameters.

1.5.5 Chapter 6

Chapter 6 presents a novel statistical method for definition of the levels and limits of normal ageing whole brain volume changes. Mean differences across age, commonly computed in neuroimaging studies, did not well represent differences at the limits of normal ageing brain volumes (errors = –74% to 75%).

1.5.6 Chapter 7

Chapter 7 assesses the ability of the Gaussian distribution to adequately model volumes from regions across the brain. The limits of brain structure and effect sizes of age were often vastly different between parametric and nonparametric methods. The assumptions of parametric methods were not met in these data which suggests that nonparametric methods are required to adequately define the normal limits of regional brain volumes.

1.5.7 Chapter 8

Chapter 8 presents novel voxel–based statistics for quantifying brain structure changes in individuals and groups. I developed software to compute voxel–wise kurtosis, skewness, and parametric and nonparametric effect sizes. Voxel values were often not distributed as assumed by parametric tests. This meant that these tests often overestimated some and underestimated other differences between normal and AD subjects. Further, I developed voxel–based brain ranking that can be used to transform voxel–values so that they are more amenable to subsequent nonparametric testing. This may also be used in the future to support diagnoses of neurological disorders such as AD and schizophrenia.

1.5.8 Chapter 9

Chapter 9 describes the development and potential uses of the Brain Images of Normal Subjects (BRAINS) databank and atlases. The BRAINS bank is a unique collection of cross-sectional and longitudinal brain images from across the lifecourse with extensive demographic, medical, and cognitive data. The BRAINS atlases were developed using novel nonparametric methods and may be used to indicate pathology in groups and individuals with diseases such as Alzheimer's.

1.5.9 Chapter 10

Chapter 10 summaries and discusses the main findings from this work. The limitations and potential future directions are also discussed.
These conclusions were drawn from interpretation of my own findings in view of existing literature. This literature is reviewed in the following three chapters.

2. Systematic review of normal ageing brain image databanks[2]

2.1 Introduction

With large volumes of data, brain MRI databanks may facilitate a better understanding of the variation in the normal ageing brain (Meehl, 1978; Cohen, 1994; Mazziotta et al., 2001; Toga, 2002; Insel et al., 2004; Toga et al., 2006; Van Horn and Toga, 2009; Freedman, 2010). For example, results from statistical models may be stored and validated with analyses of further data. These databanks and models require image data in an array range of MR sequences, such as T1, T2, T2*, and FLAIR, so to effectively describe the range of brain features in normal ageing, e.g. atrophy and white matter lesions (Wardlaw et al., 2011).

Brain MRI databanks should also address the issue of, "what is normal" (Mazziotta et al., 2009). Many older people (aged ≥60 years) without neurological disorder have clinical characteristics that may affect brain structure, e.g. hypertension, diabetes, and medication use (Jernigan et al., 2001; DeCarli et al., 2005; Grady et al., 2006; Marcus et al., 2007c; Ellis et al., 2009; Farrell et al., 2009; Mazziotta et al., 2009; Marcus et al., 2010). This suggests that these clinical characteristics should therefore be considered as part of "effects of normal ageing". Subjects with them, if otherwise cognitively normal, should probably not be excluded from normal ageing brain image studies and databanks. Instead, it should be possible to say in which way subjects are normal, i.e. whether they are ageing "successfully" (without these clinical characteristics) or "usually" (with these characteristics) (Barkhof et al., 2011). The inclusion of subjects in brain MRI studies and databanks should therefore be supported by thorough cognitive and medical test results (metadata) (Mazziotta et al., 2009). This is particularly true if these subjects are to be used as controls in studies of dementia and related disorders where cognitive state greatly influences diagnoses (Folstein et al., 1975; Morris, 1993).

If brain MRI databanks meet these prerequisites then they have the potential to support better understanding of normal ageing brain structure and diagnoses of age-

[2] Some of the work described in this chapter was published as follows:

Dickie, D.A., Job, D.E., Poole, I., Ahearn, T.S., Staff, R.T., Murray, A.D., Wardlaw, J.M. (2012). Do brain image databanks support understanding of normal ageing brain structure? A systematic review. *Eur. Radiol.* 22, 1385–1394.

related neurological disorders (Van Horn and Toga, 2009). A systematic review may determine if they have yet realised this potential and if not, what still needs to be done.

2.2 Methods

The Preferred Reporting Items for Systematic Reviews and Meta–Analyses (PRISMA) statement (Moher et al., 2009a, b), a checklist for preparing clear and transparent accounts of systematic reviews (EQUATOR Network, 2011), was used to conduct the systematic review.

Between October 2010 and October 2011, a literature search was performed using PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and the internet using Google (http://www.google.co.uk/) and Google Scholar (http://scholar.google.co.uk/) with the terms: "Magnetic Resonance Imaging" or "MRI" or "MR" and "brain" and "databank" or "database" or "data set" or "atlas" or "template" and "human". The internet search ended after two consecutive result pages provided no reference to brain image databanks.

The search was supplemented by consulting the Biomedical Informatics Research Network (BIRN; http://www.birncommunity.org/resources/data/), repositories of neuroimaging resources (http://neuinfo.org/; http://www.nitrc.org/) and reference lists in previous commentaries of brain image databanks (Mazziotta et al., 2001; Toga, 2002; Toga et al., 2006; Van Horn and Toga, 2009).

First, the abstracts and/or titles produced from the search were read to select publications potentially describing a human brain image databank. Databanks described in these selected publications were included for review if they: 1) provided publicly accessible and downloadable brain images, 2) included people aged 60 years and over, 3) described some or all of their subjects as "normal", 4) stored structural brain images of individual subjects. But not if they stored only brain atlases or templates, i.e. averaged or combined brain images from multiple subjects, without the underlying individual subjects' images. While the latter three criteria are self–explanatory, public accessibility may lead to better understanding of normal brain ageing because it allows for the sharing of data and results derived from these data (Van Horn and Toga, 2009; Freedman, 2010).

Where available, each databank was described by its: 1) purpose, 2) number of subjects, 3) number of "normal" subjects aged 60 years and over, 4) criteria for normality, 5) MR image sequences, 6) image (subject) retrieval parameters and 7) results from statistical analyses, e.g. the mean and variance of brain volumes by age, on the data contained. Missing values were estimated where possible; the bases for estimates are noted in the results tables. Finally, databanks were accessed first hand and/or their manuals and publications consulted to determine how data were accessed and searched. Tables summarising these results and individual descriptions of each databank were composed.

2.3 Results

2.3.1 Systematic search results

The literature search produced 591 publications and a further 31 items were found through internet and supplementary searches. Seven records were duplicates therefore the search produced 615 individual records. These were screened to identify 144 that potentially described a databank of human brain images. Based on criteria given in the methods section, 135 of these were excluded (Figure 5).

Particular records that did not meet inclusion criteria and the main reasons for exclusion, included:

- the Cardiovascular Health Study (CHS) database: as data were only accessible to investigators associated with the CHS study (The Cardiovascular Health Study, 2010)
- the AddNeuroMed Study database: as data were not yet publicly accessible (Simmons et al., 2011; The AddNeuroMed Study, 2011)
- the BrainSCAPE database: as it was offline at the time of this review (Neuroinformatics Research Group, 2011)
- the Whole Brain Atlas (Johnson and Becker, 1999) and Neuroanatomical Database of Normal Japanese Brains (Sato et al., 2003a, b): as images were not downloadable

- the Allen Human Brain Atlas (Allen Institute for Brain Science, 2011), NIH MRI Study of Normal Brain Development database (Evans, 2006a, b), BrainWeb: Simulated MRI Volumes for Normal Brain Database (Cocosco et al., 1997; McConnell Brain Imaging Centre of the Montreal Neurological Institute, 2004), IMAGEN project (The IMAGEN Consortium, 2011), Morphometry BIRN database (Biomedical Informatics Research Network (BIRN), 2009), and Surface Management System Database (SumsDB) (Dickson et al., 2001; Van Essen Lab, 2001): as they did not include subjects aged 60 years and over
- the International Stroke Database (Sorensen and Wu, 2010), the Neuropsychiatric Imaging Research Laboratory (NIRL) Imaging Database (Neuropsychiatric Imaging Research Laboratory, 2009), 1000 Functional Connectomes Project/ International Neuroimaging Data–sharing Initiative (INDI) database (Milham et al., 2011), Function BIRN Data Repository (Biomedical Informatics Research Network (BIRN), 2011) and Brain Image Database (BRAID) (Letovsky et al., 1998; Department of Radiology University of Pennsylvania, 2008): as they did not describe any of their subjects as "normal"
- the BrainMap database (Fox and Lancaster, 2002; Research Imaging Institute UTHSCSA, 2010), BRAINnet Database (BRAINnet Foundation, 2009), Brede Database (Technical University of Denmark Informatics, 2009), and Internet Brain Volume Database (IBVD) (The Center for Morphometric Analysis MGH HMS, 2002): as they did not contain structural brain images
- and the Montreal Neurological Institute (MNI) 152 atlas (McConnell Brain Imaging Centre of the Montreal Neurological Institute, 2010): as it did not provide the underlying individual subjects' images.

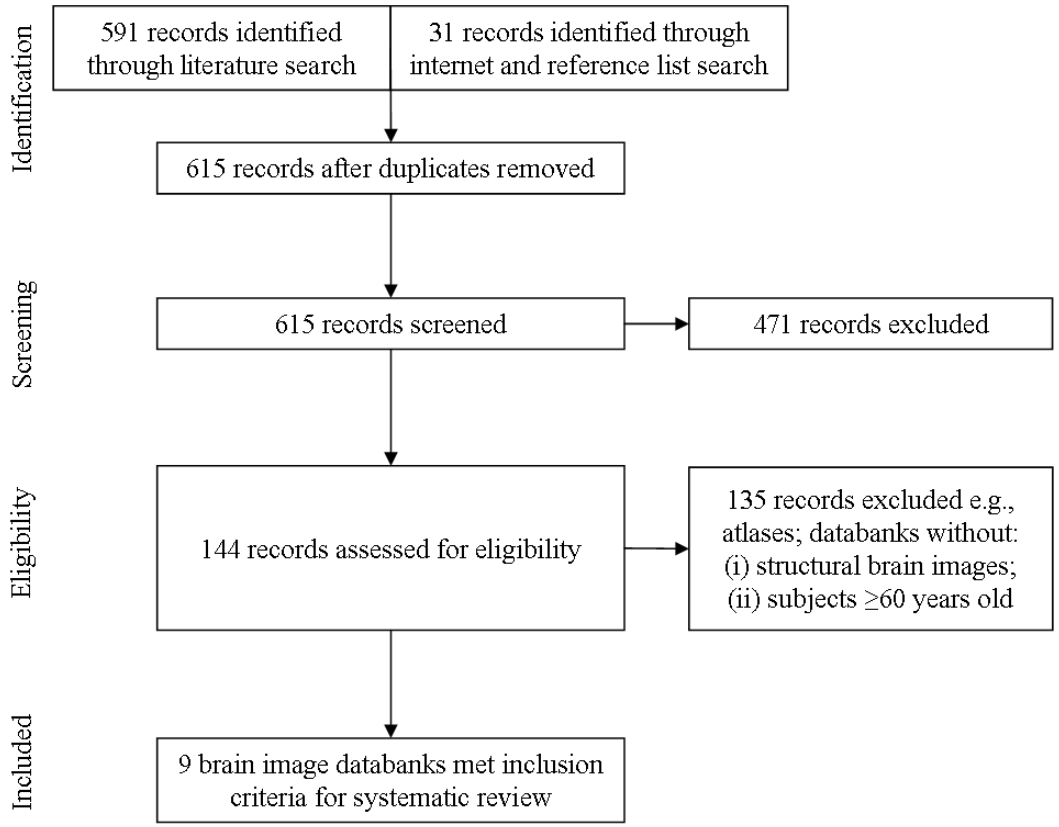


Figure 5. PRISMA flow diagram of phases in the systematic review of normal ageing brain image databanks

2.3.2 Included normal ageing brain structure MRI databanks

This left nine MR brain image databanks and 944 subjects (mean 118, SD 74 subjects per databank) that apparently met the inclusion criteria (Table 2 and Table 3). Although all of these databanks reported to include "normal" subjects aged 60 years and over, five (ADNI, fMRIDC, OASIS cross–sectional, OASIS longitudinal and XNAT Central) included subjects fully representative of the normal ageing population, according to cognitive and medical test results. Further, information on the total number of normal subjects aged ≥60 years and the criteria for normality for all subjects was not accessible in the XNAT Central databank. Furthermore, there were many subjects in more than one databank (fMRIDC, OASIS longitudinal, OASIS cross–sectional, and XNAT Central databanks) (Marcus et al., 2007c; Marcus et al., 2010).

Table 2. “Normal” ageing brain structure MRI databanks

1.	The Alzheimer's Disease Neuroimaging Initiative (ADNI) databank (Mueller et al., 2005; Jack Jr et al., 2008; Toga, 2009; Petersen et al., 2010; Laboratory of Neuro Imaging, 2011a, b)
2.	Australian Imaging Biomarkers & Lifestyle Flagship Study of Ageing (AIBL) databank (Ellis et al., 2009; Toga, 2009; Ellis et al., 2010; Laboratory of Neuro Imaging, 2011b)
3.	Designed Database of MR Brain Images of Healthy Volunteers (Mortamet et al., 2005; Bullitt et al., 2010)
4.	The fMRI Data Center (fMRIDC) (Van Horn et al., 2001, 2007)
5.	Information eXtraction from Images (IXI) dataset (Rowland et al., 2004; Biomedical Image Analysis Group Imperial College London, 2010; Hill et al., 2010)
6.	International Consortium for Brain Mapping (ICBM) databank (Mazziotta et al., 2001; Mazziotta et al., 2009; Toga, 2009; Laboratory of Neuro Imaging, 2011b)
7.	Open Access Structural Imaging Series (OASIS): Cross–sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults (Marcus et al., 2007c, b)
8.	OASIS: Longitudinal MRI Data in Nondemented and Demented Older Adults (Marcus et al., 2007b; Marcus et al., 2010)[23, 76]
9.	The Extensible Neuroimaging Archive Toolkit (XNAT) Central (Marcus et al., 2007a, 2011)

Therefore in total, according to closer inspection of available information, there were approximately 343 different individual, representative normal subjects aged ≥60 years (from the ADNI, fMRIDC and OASIS Cross–sectional databanks; Table 3). This is approximately one third of the apparent 944 subjects available on initial inspection.

Table 3. Description of "normal" ageing brain structure MRI databanks

	No. of "normal" subjects				Age range; mean, SD in years[a]	Representative cognitive and medical test results[b]	Accessibility
	60–69	70–79	≥80 years	(Total)			
ADNI databank	0	166	63	(229)	60–90; 75.8, 5.0	Yes	By application
AIBL databank	85	69	23	(177)[c]	60–100; 70.0, 7.0	No	By application
Designed Databank of MR Brain Images of Healthy Volunteers	14	1	0	(15)	60–72; 64.9, 3.1	No	Open
The fMRIDC	.	.	.	(66)[d]	60–93; 75.6, 6.7[e]	Yes	By application
IXI dataset	129	58	10	(197)	60–86; 68.1, 6.0	No	Open
ICBM databank	47	19	10	(76)[f]	60–90; ., .	No	By application
OASIS Cross–sectional databank	25	35	38	(98)[d]	60–94; 75.9, 9.0	Yes	Open
OASIS Longitudinal databank	23	35	28	(86)[d]	60–93; 75.8, 8.2	Yes	Open
XNAT Central	.	.	.	(.)[d]	60–94[g]; ., .	Yes	Open/ by application[h]

Note: .=missing value; [a]This column shows the age range; mean, and standard deviation (SD) of "normal" subjects aged ≥60 years; [b]This shows whether or not I found cognitive and medical test results, representative of the entire normal ageing population, in the databank/ criteria for normality; [c]Estimated from age frequency distribution graph (Ellis et al., 2009); [d]Many of these subjects were part of the fMRIDC, OASIS Cross–sectional, OASIS Longitudinal and XNAT Central databanks (Marcus et al., 2007c; Marcus et al., 2010); [e]Estimated from ages and sample sizes in original studies (Head et al., 2005; Grady et al., 2006); [f]Estimated from exclusion by age group graph (Mazziotta et al., 2009); [g]This is the age range according to the information found; [h]Accessibility was data dependent.

Of the 343 different individual, representative normal subjects, approximately 25 were aged 60–69, 201 aged 70–79, and 101 aged ≥80 years (I did not find the age distribution of the different individual subjects in the fMRIDC, *n*=16). The mean age of the 343 different individual, representative normal subjects was 75.78, standard deviation (SD) 6.49 years.

Many subjects were not accessible without application; only 98 different individual, representative normal older subjects were openly accessible. Databanks that were not openly accessible (Table 3) had reasonable application procedures. Applications required a description of the intended analysis and the investigators involved. This is not overly restrictive but anonymous review (similar to publication peer-review) was not adopted by any databank.

The 98 openly accessible subjects represented approximately 10% of the apparent 944 subjects available on initial inspection. The mean age of the openly accessible subjects was 75.92, SD 8.99 years. Discounting whether representative metadata were available and subject overlap, the apparent 944 subjects available (in the seven databanks where age could be found or estimated) had a mean age of 72.31, SD 6.36 years.

2.3.3 Image retrieval and acquisition parameters

Table 4 shows the subject (image) retrieval parameters and Table 5 the image acquisition parameters in the nine eligible databanks.

Subjects were retrievable by measures of whole brain and regional structures in OASIS and XNAT. No databank supported subject retrieval by more subtle confounding factors, such as Fazekas rating or quantitative volume of white matter lesions. No databank supported random or stratified random, i.e. random within specified constraints such as age, subject retrieval. Finally, no databank stored results from statistical models of the data it contained.

Detailed descriptions of each databank included in the review are in Appendix AI.1, “Individual descriptions of each databank included in the systematic review”.

Table 4. Subject retrieval parameters in “normal” ageing brain structure MRI databanks

	Clinico–demographics	Cognitive test results	MR image acquisition parameters	Image analysis results
ADNI	1	1	1	1
AIBL	1	1	1	1
Designed Database of MR Brain Images of Healthy Volunteers	1	0	1	0
The fMRIDC	0	0	0	0
IXI dataset	1	0	1	0
ICBM	1	1	1	1
OASIS Cross–sectional	1	1	1	1
OASIS Longitudinal	1	1	1	1
XNAT Central	1	1	1	1

Note: 0=were not image retrieval parameters; 1=were image retrieval parameters.

Table 5. Image acquisition parameters in "normal" ageing brain structure MRI databanks

	No. of scanning centres	T1	T2	PD	FLAIR	T2*	SWI	Tesla	T1 voxel size
ADNI databank	58	1	1	1	0	0	0	1.5, 3	1×1×1.2 mm
AIBL databank	2	1	1	1	1	0	0	3	1×1×1.2 mm
Designed Databank of MR Brain Images of Healthy Volunteers	1	1	1	0	0	0	0	3	1×1×1 mm
The fMRIDC	2	1	0	0	0	0	0	1.5	. × . ×1.4 mm
IXI dataset	3	1	1	1	0	0	0	1.5, 3	0.94×0.94×1.2 mm
ICBM databank	8	1	1	1	0	0	0	1.5, 2, 3	1×1×1 mm
OASIS Cross–sectional databank	1	1	0	0	0	0	0	1.5	1×1×1.25 and 1×1×1 mm
OASIS Longitudinal databank	1	1	0	0	0	0	0	1.5	1×1×1.25 and 1×1×1 mm
XNAT Central	1	1	0	0	0	0	0	1.5	1×1×1.25 and 1×1×1 mm

Note: these parameters are for normal ageing subjects that I could access; No.=number; PD=Proton Density; FLAIR=Fluid Attenuated Inversion Recovery; SWI= Susceptibility–Weighted Imaging; 0=image sequence was not in databank; 1=image sequence was in databank; .=missing value.

2.4 Discussion

A systemic review found 9 databanks with structural MR brain images of "normal" older people (aged ≥60 years). Amongst these databanks there appeared to be 944 normal subjects aged 60 years and over. However, many subjects were in more than one databank or did not have metadata (cognitive and medical test results) representative of the entire normal ageing population. When adjusting for these subjects, there were approximately 25 normal subjects aged 60–69, 201 aged 70–79, and 101 aged ≥80 years (I did not find the age distribution of the different individual subjects in the fMRIDC, *n*=16). In all, brain MRI from only 98 normal subjects ≥60 years were openly accessible worldwide.

There is high variation in structure of the normal ageing brain (Manolio et al., 1994; Resnick et al., 2003; Farrell et al., 2009) and inconsistency of causal inferences between studies (Good et al., 2001; Sowell et al., 2003; Fotenos et al., 2005; Salthouse, 2011). These studies had upwards of 400 subjects and their inconsistencies would suggest that 343 subjects (most of which some investigators may not be able to access) may be too few to effectively characterise normal brain structure within and between older age ranges.

The criteria for normality in many of these databanks may not have fully represented the clinical characteristics of normal ageing. For example, the Designed Database of MR Brain Images of Healthy Volunteers and ICBM criteria for normality were the same across the lifespan (Mazziotta et al., 2009; Bullitt et al., 2010). These criteria excluded the increasing proportion of prescription medications and hypertension in cognitively normal old age. When a population has a mean of approximately three prescriptions and ~50% have medically diagnosed hypertension (DeCarli et al., 2005; Marcus et al., 2007c; Ellis et al., 2009; Farrell et al., 2009; Mazziotta et al., 2009; Marcus et al., 2010), it would be reasonable to conclude that subjects with these characteristics are at least equally "normal" to subjects without. It is arguable then that both could be included in normal ageing brain image databanks. The entire normal ageing population that includes "successful" and "usual" ageing individuals would then be represented (Barkhof et al., 2011). This is particularly important when considering that these subjects may be used as controls for study groups, e.g. dementia subjects, that have similar clinical characteristics (Marcus et al., 2010).

Brain MRI databanks with representative normal older subjects have led to many publications: over 200 publications from the ADNI databank and the OASIS cross–sectional databank has been cited over 100 times (ADNI, 2011; Google Scholar, 2011). These databanks provided a limited range of image sequences (Jack Jr et al., 2008); the openly accessible images from representative normal older subjects were only T1–weighted (Marcus et al., 2007c). To support better understanding of normal ageing brain structure, databanks should include, for example, T2* and FLAIR images as well as the commonly used T1 and T2 images (Marcus et al., 2007c; Jack Jr et al., 2008; Wardlaw et al., 2011). These extra image sequences will allow for reliable identification of common brain ageing features, e.g. white matter lesions (Wardlaw et al., 2011).

Almost all databanks supported image (subject) retrieval by clinico–demographic and scanning acquisition parameters. Few (LONI IDA, OASIS, and XNAT Central) supported subject retrieval directly by medical and cognitive test results and parameters derived from image analyses e.g. brain volumes. In particular, I did not find any databank to support subject retrieval by white matter changes that, if prevalent in a normal ageing control group, could skew cognitive measures (Breteler et al., 1994). It may then be difficult to match or differentiate databank subjects (to study groups) on potentially confounding variables, e.g. cognitive scores, head size, blood pressure. Moreover, no databank supported random subject retrieval and so manual selection bias errors may occur in studies that use these databank subjects as controls.

Databanks that store results from statistical models of normal brain images and metadata may facilitate better understanding of normal ageing brain structure (Meehl, 1978; Cohen, 1994; Farrell et al., 2009; Van Horn and Toga, 2009; Freedman, 2010). For example, results from statistical models can be stored, tested, and validated with analyses of further data. The storage of statistical models is required due to the unreliability of causal inferences from single studies (Meehl, 1978; Cohen, 1994; Freedman, 2010). This is particularly true in brain imaging, given the inconsistencies in reports of normal ageing brain atrophy progression and cognition (Good et al., 2001; Sowell et al., 2003; Fotenos et al., 2005; Salthouse, 2011). Despite their need, I found no databank that had results from statistical models.

Although not yet included, the ICBM databank plans to include probabilistic atlases of the variation of normal brain structures from 7000 subjects throughout the

adult lifespan (Mazziotta et al., 2001). According to a more recent progress report of this work (Mazziotta et al., 2009), the number of subjects aged ≥60 years will be limited. From subject proportions in the progress report, I estimated 581 older subjects, if the 7000 subject target is reached. The Internet Brain Volume Database (IBVD) (The Center for Morphometric Analysis MGH HMS, 2002), excluded from my review as it did not contain structural brain images, contained brain volume measurements from different studies. These studies had different normality criteria and other parameters so may not effectively describe normal ageing brain structure variation.

This shows that there is a current lack of easily accessible, and reliable statistical models and atlases of the normal ageing brain, e.g. distributions of normal ageing brain volumes. These methods may support better understanding of normal ageing brain structure and diagnoses/ treatment of dementia (Meehl, 1978; Cohen, 1994; Selkoe, 2001; Farrell et al., 2009; Freedman, 2010). For example, dementia may be predicted and treatments dosed according to patient positions in the distribution of normal ageing brain volumes.

To have such a great effect, these models and atlases should be openly available to others (Van Horn and Toga, 2009). I conducted a formal review to determine what brain image atlases were openly available.

3. Review of magnetic resonance brain image atlases

3.1 Introduction

The basic concept of brain atlases is to create image-based models of the brain and quantify relationships between these models and a large number of variables of interest, e.g. skull structures, age, functional response. Large amounts of data now storable in brain MRI banks mean that, although even in relatively modern times they were derived from one or few subjects (Talairach and Tournoux, 1988; Ono et al., 1990; Duvernoy, 1999), brain atlases may now be derived computationally from many subjects (Evans et al., 1993). While these computational methods have significantly advanced in the last fifteen years, the brain atlas concept has been around for several hundred years (Broca and Brabrook, 1881; Anderson and Makins, 1889; Brodmann, 1994; O'Connor, 2003; Toga et al., 2006).

Modern (MRI) brain atlases[3] not only allow the mapping of various structures and quantification of relationships but also provide a standard framework, known as "*stereotactic space*", in which to compare subjects at the voxel level (Talairach and Tournoux, 1988; Evans et al., 1993; Evans et al., 2012). Standard stereotactic space is basically a set of 3D axes (X, Y, and Z) that, at the very least, allow brain MRI to be overlaid at their centres. The development of a standard space was required to sensibly compare subject voxels because head size, position, and angle vary dramatically between brain MRI subjects. In general, these variables are not known to be a direct consequence of age (after full development) or disease (Courchesne et al., 2000; Good et al., 2001; Ge et al., 2002; Buckner et al., 2004; Kruggel, 2006) and so must be controlled for in subsequent analyses. For example, without controlling for these it would be difficult to determine if reduced grey matter volume was due to age and/or disease or simply due to a person being smaller.

One of the first stereotactic atlases was developed by Talairach and Tournoux in 1988. This was derived from the brain in a 60 year old female cadaver. One

[3]Note that "MRI brain atlases" are sometimes distinguished from "MRI brain templates" in the literature. That is, where a "template" may consist of brain structure derived from one or many individuals in a standard framework; an atlas, while still consisting of brain structure derived from one or many individuals in a standard framework, also consists of labels that have been added to specific regions to identify them as, for example, hippocampal grey matter or lateral ventricle cerebrospinal fluid. However, others do not make this distinction between the terms and use them interchangeably. In this work we too use these terms interchangeably, generally preferring the term "atlas".

hemisphere of the brain was sectioned in the sagittal plane and the other in the coronal plane and from these the axial images were derived, ignoring hemispheric asymmetry (Talairach and Tournoux, 1988; Evans et al., 2012). These images were digitised to generate the electronic, "Talairach space". Although initially used as a standard framework in which to register brain MRI, several limitations of Talairach space were apparent. Ignored in the Talairach space, brains normally exhibit hemispheric asymmetry and as it was only based on one subject, the Talairach brain did not represent the great variance in brain structure among individuals. These limitations are clear, have been discussed extensively (Mazziotta et al., 2001), and have been well known since the atlases' inception (Evans et al., 1993). Indeed, they were noted in the original author's foreword, "this method is valid with precision only for the brain under consideration" (Talairach and Tournoux, 1988).

Given the limitations inherent in single subject atlases, the MNI created an atlas from 305 brain MRI subjects in the early 1990s (Evans et al., 1993). These 305 subjects were initially registered to the Talairach space by aligning their manually identified anterior–posterior commissure (AC–PC) line to the Talairach AC–PC line and adjusting the extent of their XYZ axes to the extent of the Talairach XYZ axes (Figure 6). The average voxel–wise MRI intensity of the subjects was then derived in Talairach space to create "initial MNI space". The subjects were then all re–registered to initial MNI space and a final average was taken to create "final MNI space", or just "MNI space" as it is known in the literature.

The MNI305 atlas circumvented the single subject bias inherent in the Talairach atlas and, through its apparent loss of resolution (it was rather blurry in some regions), somewhat reflected the variance in brain structure between a large number of subjects (Evans et al., 1993). However, it did not actually define any measures of voxel variance, e.g. SD or interquartile range (IQR), therefore reflected the central tendency but not the range of normal brain structure. Further, the lack of complete agreement between the shape of each of the 305 subjects and the shape of the Talairach brain (bottom panel in Figure 6) led to the MNI atlas being markedly larger than the brain of an average person (although this is potentially no worse than perfectly matching the size of the Talairach brain which was smaller than average). Furthermore, while not based on the symmetrical sulci pattern of Talairach, the MNI atlas was derived in the same general (symmetrical) space therefore could lead to bias

towards symmetrical hemispheric width – something that does not often occur naturally (Raz et al., 2004).

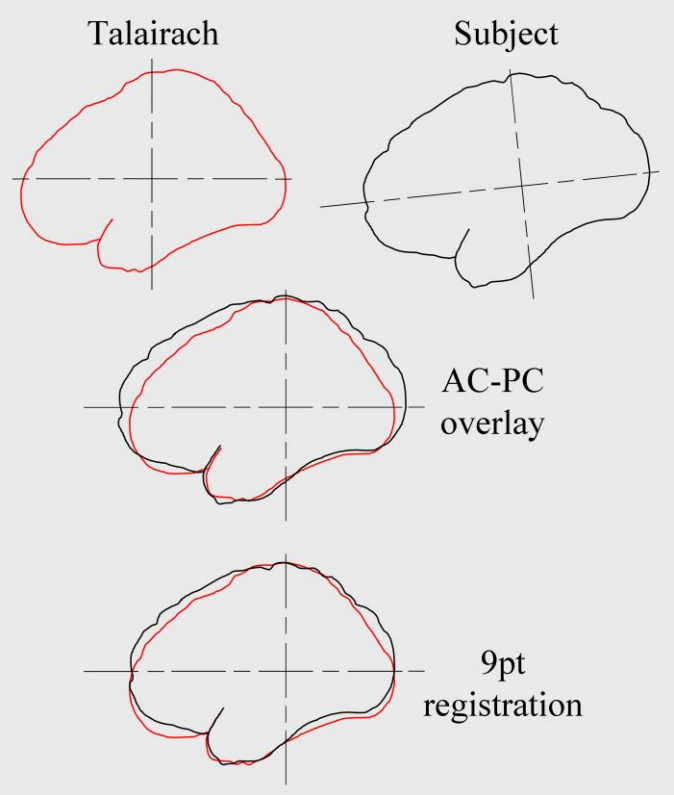


Figure 6. Schematic of "9pt" registration from a subjects' native space (brain size and position) into the size and position of the Talairach brain
This method was used to derive the first generation MNI atlas, "MNI305". "9pt" refers to the overlay, alignment, and stretching of the 3 axes, X, Y, and Z. AC–PC= anterior–posterior commissure.

Despite these potential issues, it was apparent from a brief review that the MNI atlas (also referred to as the International Consortium for Brain Mapping (ICBM) atlas) was one of only a few atlases in wide use (Ashburner and Friston, 2000; Good et al., 2001; Mazziotta et al., 2001; Evans et al., 2012). Although later generations of the MNI atlas were released, e.g. MNI152, these were all derived from mainly young adult subjects. Brain structure in old age is different to brain structure in younger and middle–aged adults (Gur et al., 1991; Courchesne et al., 2000; Good et al., 2001; Sowell et al., 2003). For example, the lateral ventricles and sulci spaces are generally larger in subjects over 60 years (Lemaitre et al., 2005). Because of these differences in

brain structure the use of an atlas based on only younger subjects can create a bias in lifecourse studies, e.g. systematic overexpansion of older brains (Buckner et al., 2004). An atlas based on only aged subjects would be appropriate for a study involving only aged subjects but if implementing two atlases (one for younger subjects and another for older subjects) in a lifecourse study, differences between ages could be attributed to differences between the atlases.

There is therefore a need for a range of brain atlases for different age ranges and I undertook a formal review to determine if these already existed.

3.2 Methods

I conducted a brain MRI atlas search concurrently with the systematic review of brain MRI databanks (chapter 2). However, the atlas review described in this chapter was not systematic and did not follow PRISMA guidelines due to a lack of time. Between October 2010 and October 2011 I searched PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and the internet using Google (http://www.google.co.uk/) and Google Scholar (http://scholar.google.co.uk/) with the terms: "Magnetic Resonance Imaging" or "MRI" or "MR" and "brain" and "atlas" or "template" and "human". As noted in the introduction to this chapter and section 1.1.4, the term "atlas" is sometimes distinguished from the term "template" in the literature. That is, where a "template" may consist of brain structure derived from one or many individuals in a standard framework; an atlas, while still consisting of brain structure derived from one or many individuals in a standard framework, also consists of labels that have been added to specific regions to identify them as, for example, hippocampal grey matter or lateral ventricle cerebrospinal fluid. However, others do not make this distinction between the terms and use them interchangeably. Not to exclude records simply due to personal nomenclature preference, I searched for both "atlases" and "templates" (and use these terms interchangeably, preferring the term, "atlas").

I supplemented this search until August 2013 by periodically searching Google with a subset of these terms and by reviewing content alerts distributed by relevant journal articles, e.g. NeuroImage (http://www.journals.elsevier.com/neuroimage/) and Human Brain Mapping (http://onlinelibrary.wiley.com/journal/

10.1002/(ISSN)1097–0193). I also reviewed the references in previous brain atlas commentaries (e.g. Mazziotta et al., 2001; Toga et al., 2006; Evans et al., 2012).
I did not have a systematic exclusion process in this review, rather, I only excluded atlases that were apparently not publicly available, e.g. the elderly and demented sub–volume probabilistic brain atlas (that consisted of 20 subjects; Mega et al., 2005), animal brain atlases, functional brain MRI image atlases, and those that did not have adult (>18 year old) subjects (I did not maintain a record of all the atlases I excluded). The next section therefore describes all structural brain MRI atlases that I found to be publicly available and include adult human subjects.

3.3 Results

3.3.1 Structural brain MRI atlases

The 31 structural brain MRI atlases that I found to be publicly available and include adult subjects are listed in Table 6. These atlases are described in more detail in section 3.3.3; the subjects used to create each of them are described in Table 7.

3.3.2 Subjects in structural brain MRI atlases

The subjects used to create each of the structural brain MRI atlases are described in Table 7. The "Digital Anatomist: Interactive Brain Atlas" and "Whole Brain Atlas" are not listed because they did not provide individual subject information (Sundsten, 1994; Johnson and Becker, 1999). The 30 atlases (including the two parts of "Normal reference MR images for the brain" separately) that provided individual subject information had between 1 and 662 subjects (median 27, IQR 44.5). With the exception of the "Alzheimer's Disease Template", "Normal reference MR images for the brain", "Brain atlas for healthy elderly", "Symmetric atlas in normal older adults" and "Clinical toolbox", most atlases were derived from mainly young to middle aged adults. Three atlases spanned the majority of adult life (18–90 years), both "FreeSurfer" atlases and the "SRI24" atlas, of which the largest had 40 subjects. There were three atlases for focused age ranges, both "Normal reference MR images for the brain" (65–70 and 75–80 year old subjects), and the "PALS" atlas (18–24 year old subjects).

Table 6. Structural brain MRI atlases

1. Allen Human Brain Atlas (Allen Institute for Brain Science, 2011)
2. Alzheimer's Disease Template (Thompson et al., 2000; Thompson et al., 2001)
3. Brain atlas for healthy elderly (Lemaitre et al., 2005)
4. Chinese_56 (Tang et al., 2010)
5. Clinical toolbox (Rorden et al., 2012)
6. Digital Anatomist: Interactive Brain Atlas (Sundsten, 1994)
7. FreeSurfer 'Desikan–Killiany' cortical atlas (Desikan et al., 2006; Schmansky, 2010)
8. FreeSurfer 'Destrieux' cortical atlas (Fischl et al., 2002; Fischl et al., 2004; Schmansky, 2010)
9. Harvard brain atlas (Shenton et al., 1995)
10. Harvard–Oxford cortical and subcortical structural atlases (FMRIB, 2008)
11. MNI/ ICBM 152 (Mazziotta et al., 2001)
12. MNI/ ICBM 452 (Mazziotta et al., 2001)
13. JHU ICBM–DTI–81 white–matter labels atlas (Mori et al., 2005; FMRIB, 2008)
14. JHU white–matter tractography atlas (FMRIB, 2008; Hua et al., 2008)
15. Jülich histological atlas(Eickhoff et al., 2005; FMRIB, 2008)
16. LPBA40 (Shattuck et al., 2008)
17. MNI single–subject brain ("colin27")(Collins et al., 1998; Holmes et al., 1998)
18. MNI structural atlas (Collins et al., 1995; Mazziotta et al., 2001; FMRIB, 2008)
19. MNI305 (Evans et al., 1993)
20. Normal reference MR images for the brain (Farrell et al., 2009)
21. Oxford thalamic connectivity atlas (Johansen-Berg et al., 2005; FMRIB, 2008)
22. Population–average, landmark–and surface–based (PALS) atlas (Van Essen, 2005)
23. Parcellated MNI single–subject brain (Tzourio-Mazoyer et al., 2002)
24. SRI24 (Rohlfing et al., 2010)
25. Symmetric atlas in normal older adults (Grabner et al., 2006; Zamboni et al., 2013)
26. Talairach Daemon (Lancaster et al., 1997; Lancaster et al., 2000)
27. Talairach and Tournoux (Talairach and Tournoux, 1988; Brett et al., 2001; Brett et al., 2002)
28. The Cerefy clinical brain atlas (Nowinski and Belov, 2003)
29. The Whole Brain Atlas (Johnson and Becker, 1999)
30. Washington University 711 target atlases (Buckner et al., 2004)
31. WFU_PickAtlas (Maldjian et al., 2003; Maldjian et al., 2010)

Note: ICBM=International Consortium for Brain Mapping; JHU=Johns Hopkins University; DTI=Diffusion tensor imaging; LPBA= LONI Probabilistic Brain Atlas; MNI=Montreal Neurological Institute; SRI=Stanford Research Institute; WFU=Wake Forest University.

Table 7. Subjects used to create structural brain MRI atlases

Atlas	Age in years[1]	*n* Gender[2]
Allen Human Brain Atlas	24–57; 43, 12[P]	2 F, 6 M (8)
Alzheimer's Disease Template[T]	.–.; 75, 2	14 F, 12 M (26)
Brain atlas for healthy elderly[T]	63–75; 70, 3	331 F 331 M (662)
Chinese_56[T]	21–29; 24, 2	0 F, 56 M (56)
Clinical toolbox[T]	.–.; 73, 8	32 F, 18 M (50)
FreeSurfer 'Desikan–Killiany' cortical atlas	19–86; 56, .	26 F, 14 M (40)
FreeSurfer 'Destrieux' cortical atlas	23–68; ., .	. F, . M (12)
Harvard brain atlas	25	0 F, 1 M (1)
Harvard–Oxford cortical and subcortical structural	18–50; ., .	16 F, 21 M (37)
MNI/ ICBM 152[T]	18–44; 24[*], 7	66 F, 86 M (152)
MNI/ ICBM 452[T]	"young adults"	. F, . M (452)
JHU ICBM–DTI–81 white–matter labels atlas	18–59; 39, .	39 F, 42 M (81)
JHU white–matter tractography atlas	.–.; 29, 8	11 F, 17 M (28)
Jülich histological atlas	"post–mortem brains"	. F, . M (10)
LPBA40[T]	19–39; 29, 6	20 F, 20 M (40)
MNI single–subject brain ("colin27") [T]	~35	0 F, 1 M (1)
MNI structural atlas[T]	.–.; 29, 3	. F, . M (>50)
MNI305[T]	.–.; 23, 4	66 F, 239 M (305)
Normal reference MR images for the brain[I/II]	65–70; 67, .	0 F, 54 M (54)
	75–80; 77, .	7 M, 18 F (25)
Oxford thalamic connectivity atlas	.–.; 35, 11	4 F, 7 M (11)
PALS	18–24; ., .	6 F, 6 M (12)
Parcellated MNI single–subject brain[T]	~35	0 F, 1 M (1)
SRI24	19–84; 52, 5	12 F, 12 M (24)
Symmetric atlas in normal older adults[T]	.–.; 75, 6	. F, . M (153)
Talairach Daemon[T]	60	1 F, 0 M (1)
Talairach and Tournoux	60	1 F, 0 M (1)
The Cerefy clinical brain atlas[T]	60	1 F, 0 M (1)
Washington University 711 target atlases[T]	.–.; 49, .	15 F, 9 M (24)
WFU_PickAtlas[T/MNI]	18–44; 24*, 7	66 F, 86 M (153)

Note: [1]age is displayed "*range; central tendency (median/ mean), variance (standard deviation/ interquartile range)*" when I was able to obtain it (and when there was more than one subject); [2]number of females, number of males (total); P=postmortem; T=descended from the Talairach and Tournoux atlas; .=I was unable to obtain this information; *=Median; [#]Subjects may have been schizophrenic or normal, authors "were blind to diagnosis" (Fischl et al., 2004); I/II=this consisted of two separate atlases one for each of the age groups shown; T/MNI=this atlas had two versions, one in Talairach space and the other in MNI space.

The methods used to construct these atlases and their potential uses (according to and if given by the authors) are described in the following sections.

3.3.3 Methods and uses of structural brain MRI atlases

The technical methods used in development of the atlases I found are described in the following sections. Many atlases are first based on registration of subjects into a common space. Registration, or "spatial normalisation", is the process of overlaying subjects on top of one another. It may also involve altering their brain size and shape. There are two main types of registration: linear and nonlinear. Linear registration aligns the axes of individual brains and nonlinear registration attempts to match voxel values between brains. Linear registration has between 3 and 12 parameters: 3 and 6 parameter registrations align the axes of brains and 7, 9, and 12 parameter registrations alter the axis dimensions (where the higher the number of parameters produces higher alterations compared to the original brain structure). Registration is discussed further in chapter 5 and the following sections describe the registration method (if any) used in the development of each atlas I found.

3.3.3.1 ALLEN Human Brain Atlas

Post–mortem scans were obtained from eight subjects with no brain disorder, cancer, infectious disease, or substance dependency related death. Brain images were manually labelled in their native space (i.e., substructures such as the hippocampus were delineated). Each subject contributed at least two acquisition sequences, T1 and T2, four subjects also had DTI, to provide 20 separate, structurally labelled, atlas images.

3.3.3.2 Alzheimer's Disease Template

All 26 subjects were initially registered into Talairach space before being re–registered to the sample average MRI intensity image by continuum–mechanical (highly nonlinear) registration. A final average MRI intensity image was then derived.

3.3.3.3 Brain atlas for healthy elderly

One–hundred and twenty of the 662 subjects were randomly selected and nonlinearly registered to the MNI152 template. The average tissue proportion in each voxel was then computed from these 120 subjects (in the MNI152 template space). All subjects were then nonlinearly registered to this newly created average of 120 subjects, referred to as the "Epidemiology of Vascular Aging" (EVA) prior space. The final atlas (consisting of four components) was created by averaging the smoothed T1 intensity, GM, WM, and CSF proportion images of all 662 subjects in the EVA prior space (each average image was derived separately so the atlas consisted of four separate images).

3.3.3.4 Chinese_56

A randomly selected image from the sample was used as the space to which all remaining 55 subjects were registered via a 12 parameter linear registration. An intensity average image was then derived before all images were realigned to this average by a 6 parameter linear registration. Another average image was then derived as the linear atlas. A nonlinear atlas was also created by nonlinearly registering individual images to the linear atlas and subsequently deriving the average of these voxel intensities.

3.3.3.5 Clinical toolbox

The clinical toolbox contained a CT and MRI brain atlas from 50 subjects. The MRI images were manually trimmed to remove excess signal from the neck and around the head and then linearly registered with the position and angle of MNI space. In this space, grey matter, white matter, and cerebrospinal fluid were segmented from each individual and an average atlas of each tissue was created with Diffeomorphic Anatomical Registration using Exponentiated Lie algebra (DARTEL – a function included in Statistical Parametric Mapping; Ashburner, 2007; Ashburner and Friston, 2009). These atlases and the original T1 images were then registered into MNI space with SPM's linear registration. Finally, to make the atlases symmetrical, the hemispheres within each atlas were averaged. The clinical toolbox was intended to provide a more appropriate atlas for segmenting tissue volumes in older adults.

3.3.3.6 Digital Anatomist: Interactive Brain Atlas

This atlas appeared to be an online teaching tool that consisted of several parts: (1) a cadaver brain surface; (2) cadaver brain dissections; (3, 4) cadaver brain axial and coronal slices; (5, 6, 7) brain MRI axial, coronal, and sagittal slices; (8, 9) cadaver brainstem and sections; (10) cadaver spinal cord slices; (11) sectioned blood vessels; and (12–16) a series of 2D and 3D animations.

3.3.3.7 FreeSurfer 'Desikan–Killiany' cortical atlas

Thirty–four gyral based structures were manually delineated in 40 subjects and integrated using the FreeSurfer 'Destrieux' cortical atlas method described below. Like the 'Destrieux' atlas, the 'Desikan–Killiany' atlas was proposed for automatic labelling of brain structures in new subjects.

3.3.3.8 FreeSurfer 'Destrieux' cortical atlas

Two images for each of the 12 subjects were averaged and manually segmented to label each voxel as belonging to one of 85 brain structures. Individual cortical surfaces were then reconstructed and combined into a spherical cortical atlas by landmark driven registration (manually determined points in the brain were matched between subjects rather than axis or voxels). This provided an image with probabilities of all anatomical labels at every point in the cortex given constraints of anatomical arrangement, e.g. the amygdala generally resides anterior and superior to the hippocampus but never inferiorly.

This atlas parcellates new images by assigning the highest probability label to individual voxels based on their coordinate in the atlas space and the labelling of 6 neighbouring voxels in all of the cardinal directions. The algorithm is iterated to maximise the label probability at each voxel and ends when no voxel probabilities change through an iteration. Finally, voxels are transferred from the spherical atlas space back to their native space with the label they received in spherical atlas space.

3.3.3.9 Harvard brain atlas

The single subject used in this atlas was selected from 15 others due to it showing the best tissue contrast and least artefact. The image was segmented into different tissue

classes and ROI (i.e., substructures such as the hippocampus) were drawn. Volumes and surface models were computed for each structure.

3.3.3.10 Harvard–Oxford cortical and subcortical structural atlases

Using locally sourced semi–automatic tools, each of the 37 subjects were individually segmented into 48 cortical and 21 subcortical structures. All images were then linear registered to MNI152 space before the transforms were applied to the individual labels. Finally, all images were overlaid in this space to form probability maps for each structure.

3.3.3.11 MNI/ ICBM152

All 152 subject images were individually registered to the MNI305 template using a 9 parameter linear transformation. The average intensity image of these 152 subjects was then derived in MNI305 space so to create MNI152 space.

3.3.3.12 MNI/ ICBM452

The 452 subject images were individually registered to the MNI305 template using a 12 parameter linear registration. An average image was then derived from the images in this space to create the linear atlas. To create a nonlinear atlas, the 452 subjects were individually registered to the MNI305 template using a 5 order polynomial nonlinear registration. The voxel–wise average of these nonlinearly registered images was then derived.

3.3.3.13 JHU ICBM–DTI–81 white–matter labels atlas

The 81 subjects used to create this atlas were provided by the ICBM DTI workgroup and each subject was hand segmented into 50 white matter tract labels. The result was an ICBM space average of diffusion MRI tensor maps from 81 subjects.

3.3.3.14 JHU white–matter tractography atlas

Eleven white matter structures were probabilistically defined by overlaying the fibre track ROI drawings of each of the 28 subjects in standard space.

3.3.3.15 Jülich histological (cyto– and myelo–architectonic) atlas

The histological volumes of these 10 brains were digitally reconstructed and spatially normalised to the MNI single subject atlas to create a probabilistic map of 52 grey matter and 10 white matter structures. This map was then linearly registered into MNI152 space.

3.3.3.16 LPBA40

Each of the 40 subjects were first aligned to the MNI305 atlas by a 6 parameter registration that accounted for head tilt and alignment. Fifty–six ROI structures were then manually delineated in each of the images before three versions of the atlas were created by registering the delineated brains to three different targets. LPBA40/AIR was created by nonlinear registration and intensity averaging in the ICBM452 space; LPBA40/FLIRT by linear registration and intensity averaging in the skull–stripped ICBM152 space; and LPBA40/SPM5 by nonlinear registration and intensity averaging in the whole head ICBM152 space.

Every voxel, within each atlas version, was assigned a probability for belonging to each of the delineated structures. Maximum probabilities from every voxel generated three maximum likelihood segmented atlases. Tissue maps were generated by automatically segmenting GM, WM, and CSF in each subject's native space before registering each image to atlas space. Voxels in each tissue atlas were displayed in a gradient according to the probability that they belonged to the corresponding tissue. A maximum probability map of GM was created by calculating voxels that most frequently (among all subjects) belonged to GM. ROI probability maps were created by calculating the maximum probability label in each voxel.

3.3.3.17 MNI single–subject brain ("colin27")

A single subject was scanned 27 times with each image linearly registered to MNI305 space before an average intensity image was derived.

3.3.3.18 MNI structural atlas

A single subject image was hand segmented into 9 anatomical structural regions before the labels were propagated to more than 50 subject images using nonlinear registration. Each resulting labelled brain was then transformed into MNI152 space

using linear registration. Finally, all segmented images were overlaid in this space to produce the final probabilistic atlas.

3.3.3.19 MNI305

Each of the 305 subjects were oriented to Talairach space by manually identifying their AC–PC line and overlaying it with the Talairach AC–PC line. The extent of each subject's brain was then scaled to the extent of the Talairach brain in the XYZ axes. The oriented subjects were then averaged before being individually reoriented and scaled to this average. A final average was then derived.

3.3.3.20 Normal reference MR images for the brain

A radiologist rated all subjects individually by ventricular and cortical atrophy to designate five subjects as quartile ranks in each sample (n=54, 65-70 years; n=25, 75-80 years). This inferred the distribution of atrophy within the two subject samples. This distributional atlas consists of 4 axial image slices per quartile for each subject group and was printable from: http://www.sbirc.ed.ac.uk/documents/ageing-brain-template.pdf.

3.3.3.21 Oxford thalamic connectivity atlas

Using probabilistic diffusion tractography, the probability of anatomical connection between points in the thalamus and seven cortical zones was defined in each subject. Atlas probabilities were derived from a combination of all 11 subjects.

3.3.3.22 PALS

All 12 subjects were first registered to the Washington University 711–2C atlas (section 3.3.3.30) by a 12 parameter linear registration and averaged to create the atlas volume. The fiducial surface of each hemisphere from individual images was then registered to the atlas volume by landmark constrained registration. The average representations of landmarks in the volume space then provided this surface–based atlas.

3.3.3.23 Parcellated MNI single–subject brain

Global sulcal topography was delineated in the MNI single–subject brain before 31 sulci was identified. These sulci was used to delineate 45 volumes of interest. Voxels were then labelled by the volume in which they resided. Points of functional activation in new subjects, registered to this map, may be attributed to the volume in which they reside, to the volume which they are nearest to, or to volumes in which a 10 mm radius around the point of activation encompasses.

3.3.3.24 SRI24

The anatomical and diffusion MR images for all 24 subjects were first brought into alignment. Subjects were then all registered to an empty grid of pixels, with the same characteristics as the T1 images, by non–linear registration to the group wise configuration in this space. Tissue maps were created by extracting tissue volumes separately from individual images before averaging them in the atlas space. A consolidated tissue map was created by overlaying all tissue segmented images into the atlas space and labelling voxels as per the tissue that fell most frequently within them. In the native space of subjects diffusion tensor fields were reconstructed before they were reformatted into the atlas space by a cubic interpolation kernel and averaged. Finally, two cortical parcellation maps were constructed by transferring anatomical labels, manually delineated elsewhere (Tzourio-Mazoyer et al., 2002; Shattuck et al., 2008), to the atlas space. The manually delineated images were individually registered to SRI24 space before being combined by label fusion to create the final parcellation maps.

3.3.3.25 Symmetric atlas in normal older adults

Each of the 153 subjects were initially linear registered to the MNI152 symmetric atlas. The left and right hemispheres of each subject were averaged to make each subject symmetrical. All symmetric subjects were then averaged in this space. Each subject was then re–registered to this new average, again with linear registration, and another average was derived. This process was then repeated 9 times but with nonlinear registration before a final highly symmetrical average was derived. This atlas was designed to automatically extract subcortical structures and allow left–right comparison of the volumes of these structures without bias or subjective error.

3.3.3.26 Talairach Daemon

The (single subject) Talairach atlas was digitised into a lookup table; each voxel (coordinate) was assigned a tissue type and Brodmann area.

3.3.3.27 Talairach and Tournoux

One hemisphere of a cadaver brain was sliced in the sagittal plane and the other hemisphere was sliced in the coronal plane to create symmetrical drawings of the axial sections. The brain was orientated so that a line between the AC and PC was horizontal. When new brains are registered to Talairach space their AC–PC line is oriented to the Talairach AC–PC line. The AC is the centre of the Talairach coordinate system where 3 dimensional (XYZ) axes begin. The extent of the Talairach brain on each of these axes provides the dimensions to which new brains are scaled when registered to Talairach space.

3.3.3.28 The Cerefy clinical brain atlas

The Talairach atlas was digitised; cortical and subcortical structures were segmented and labelled.

3.3.3.29 The Whole Brain Atlas

Images from various clinical and normal cases were collected and brain structures annotated. Although there were a large number of annotated images, I was unable to obtain further provenance of the images.

3.3.3.30 Washington University 711 target atlases

Four images from each of the 24 subjects were aligned and averaged to create a single, high signal-to-noise ratio (SNR) image for each subject. Minimum mismatch averages were then separately derived for each age group by first registering images to Talairach space and then generating a new target space by the average of the registered images. Images were then re-registered to the new target space and an updated target space created by further averaging. Five iterations of this process were completed in each group before all of the fifth iteration averages (the averages from each age group) were combined to generate the final atlas image.

3.3.3.31 WFU_PickAtlas

Using the Talairach Daemon, voxels (coordinates) in Talairach and MNI space were filled with Brodmann, lobular, hemispheric, anatomic, and tissue type labels. Once registered to either space voxels of subject images are labelled corresponding to the label in the coordinate at which they lie. As suggested by the information held at each coordinate in the tables, brain images that have been registered to the WFU_PickAtlas can be segmented by Brodmann areas, the major lobes, hemispheres, anatomic structures, and/or tissue types.

3.3.4 Summary of structural brain MRI atlas methods

The overwhelming majority of atlases were created via voxel-wise averaging of original MRI intensity or tissue proportion images from a group of subjects. These atlases are valuable for image processing, i.e. segmenting tissues in new subjects, and for aligning functional imaging subjects into a common space. I only found one atlas that defined the distribution of brain structure in a subject group ("Normal reference MR images for the brain"). This was not the voxel-wise distribution but the distribution of whole brain volumes as determined by qualitative radiological assessment. There was apparently no atlas that could be used to assess new subjects at the voxel level.

3.4 Discussion

I found 31 publicly available structural brain image atlases that were designed for a wide range of purposes, e.g. to provide a standard analysis space or reference for correlating brain structure and function. These were mainly derived from the mean brain structure of young to middle aged subjects. The number of subjects was generally rather small (median 27, IQR 44.5) given that several hundreds or even thousands of subjects are required to adequately represent population brain structure (Toga, 2002; Toga et al., 2006; Evans et al., 2012). Further, many were descended from a single subject atlas, the Talairach atlas.

The Talairach brain atlas has a number of limitations that were noted in its publication (Talairach and Tournoux, 1988). Most relevant to modern atlases derived

in this space, the Talairach atlas is based on symmetrical hemispheric volumes. This means that, while their sulcal pattern may be accurately asymmetric, the widths of many modern atlases' hemispheres may be inaccurately the same. The use of a consistent standard framework, e.g. a Talairach-based coordinate system, was well founded and conceptually appropriate. However, to accurately represent brain structure in a given group, it may be more appropriate to use only the standard "skeleton", i.e. axes, of Talairach space and fill that space with the asymmetric, or otherwise distinctive, structure of that group.

Specific groups should be analysed using an atlas derived from other subjects in that population, otherwise systematic errors may be introduced, e.g. the overexpansion of atrophied brains registered to younger subject atlases (such as the MNI series of atlases; Buckner et al., 2004). I found five atlases derived from ageing (≥60 year old) subjects and only three of these ("Clinical toolbox", "Symmetric atlas in normal older adults", and "Brain atlas for healthy elderly") were designed for quantitative image processing of normal subjects, each of which were derived or mapped into Talairach/ MNI space. Three atlases were derived from subjects across the adult lifecourse (both "FreeSurfer" atlases and the "SRI24" atlas) of which the largest had 40 subjects. Given the wide variation in brain structure across the lifecourse (Good et al., 2001; Sowell et al., 2003; Allen et al., 2005; Raz et al., 2010), these studies may require atlases with many more subjects. Equally, studies of focused age groups (≤5 years) require atlases specific to them. However, I found only one atlas, "PALS" (18-24 years), for image processing in such focused studies.

All but one of the atlases ("Normal reference MR images for the brain") were primarily designed for image processing that would produce data for subsequent statistical analyses. That is, they were designed to provide a standard space for voxel-wise analyses or support tissue/ subregional, e.g. hippocampus, volume extraction. The "Normal reference MR images for the brain" atlas was based on qualitatively determined percentile ranks of normal ageing brain volumes and designed to support clinical diagnoses of whole brain volume in ageing (65-70 and 75-80 year old) patients. Given that neurological diseases may lead to only subtle changes (Shenton et al., 2001; Farrell et al., 2009; Ferguson et al., 2010), it may be useful to derive a voxel-based atlas of the clinically normal distribution of brain structure. All voxel-based atlases that I found used mean-based measures of brain structure and I found none that used nonparametric, e.g. median or percentile rank, distributional measures.

The atlases in this review were found through a formal, but not systematic search, and were openly accessible. This suggests that I may not have reviewed all relevant atlases, i.e. those described as part of larger studies (and therefore potentially only visible through a full systematic search) or those not openly accessible. Further, I did not investigate potential uses for atlases beyond those described in the original manuscripts/ sources. It could be that any one of these atlases may be modified to serve additional purposes. Related to this, I described the methods and uses of each atlas according to my own interpretation of the source manuscripts/ reference manuals, which (much like a "*Chinese whisper*") may have altered the meaning intended by the original authors. Notwithstanding these limitations, I have compiled a comprehensive review that assessed the methods and uses of the most commonly applied atlases in brain imaging analyses.

I have reviewed and described openly accessible structural brain image atlases and found that, in general, they were primarily designed to serve image processing tasks in young to middle-aged adults. I found no atlas to quantify the range (distribution) of voxel-wise brain structure. In other words, there was no atlas that calculated the distribution (rather than mean) of voxel values in a group of normal subjects to determine "how normal" each voxel value was in a new subject. Rather than this atlas-based assessment, MRI brain structure was previously assessed using statistical models and these are described in the next chapter.

4. Review of statistical models of normal brain structure across the lifecourse

4.1 Introduction

Statistical models aim to describe the association between two or more variables, e.g. age and brain tissue volume. Traditionally, there were two major schools of thought in statistical modelling; there were the "frequentists" and the "Bayesians". The frequentists took the conventional view of probability, i.e. there are two sides of a coin therefore the probability of it landing on one side, say heads, is always 0.5. While the Bayesians took the view that probabilities are not constant but change given acquired data, i.e. if a coin is tossed ten times and lands on heads 6 times, then the Bayesian probability for heads on the 11^{th} toss is somewhere between 0.5 and 0.6. These were once competing views but now statisticians are generally open to using either method or a combination of both methods, depending on the situation. In brain image analyses, the frequentist approach is the most commonly applied (Ashburner and Friston, 2007; Freedman et al., 2007).

Even within the frequentist approach there are varying methods. That is, there are "*parametric*" or "*Gaussian*" (mean–based) models versus "*nonparametric*" or "*distribution–free*" (rank–based) models (Hogg and Tanis, 2010). Many datasets from a broad range of fields, e.g. the physical and life sciences, are analysed with parametric models (Freedman et al., 2007; Freedman, 2010; Hogg and Tanis, 2010). This is appropriate because these data are often distributed similarly to the Gaussian distribution (Freedman et al., 2007; Hogg and Tanis, 2010). However, many other types of data, including brain image data (Rorden et al., 2007), are not always distributed Gaussian; in which case nonparametric models may be more appropriate (Hogg and Tanis, 2010).

Although they do not assume that data are distributed Gaussian, many nonparametric methods still make assumptions about data distributions between groups. That is, frequently applied methods such as the Wilcoxon–Mann–Whitney test and "nonparametric regression" assume that distributions between groups have similar shapes and approximately equal variance (Hollander and Wolfe, 1973; Cliff, 1993; Zimmerman, 1998). When their respective assumptions are not met, the results from parametric and nonparametric tests may be unreliable and difficult to interpret (Meehl, 1978; Cliff, 1993; Cohen, 1994; Zimmerman, 1998; Freedman et al., 2007;

Freedman, 2010). In these cases, rather than comparing central tendency (mean or median) values between groups, it may be better to model the entire distributions of data between groups, e.g. values that define where a given percentage of subjects lie within a group (percentile rank values).

Further to the distributions of data, statistical models make the assumption that data were generated via a statistically random sample from a population where each individual had an equal chance of selection (Meehl, 1978; Cohen, 1994; Freedman et al., 2007; Freedman, 2010; Hogg and Tanis, 2010). A statistically random sample is equivalent to "blindly drawing tickets from a box" where the entire contents of the box are the population and the tickets drawn are the sample (Freedman et al., 2007). These samples are often difficult to generate in social and epidemiological science studies; in which case "*effect sizes*" should be computed and compared across all available data (Meehl, 1978; Cohen, 1994; Freedman et al., 2007).

One of the most widely used measures of effect size is "*Cohen's d*"; this is essentially the mean difference between groups normalised by their pooled SD (Cohen, 1988; Coe, 2002). As it is based on the standardised mean difference, this statistic still relies on the data being equally Gaussian between groups. Even small departures from these assumptions may lead to uninterpretable and/or unreliable results (Cliff, 1993; Coe, 2002). Thankfully, there is a nonparametric measure of effect size that does not require data to be Gaussian, equally shaped, or have approximately equal variance - the so called, "*d* statistic"

The "*d* statistic" is a type of nonparametric statistic that provides a direct measure of the overlap between two distributions, while not making any assumptions about data distributions (Siegel and Castellan, 1988; Cliff, 1993; Coe, 2002). Given that it does not make these assumptions, a measure of overlap is all that it provides, i.e. it does not say anything about the central tendencies (means/ medians) of groups (Cliff, 1993). The measure of overlap is a clinically relevant and useful measure in itself. This is because, despite being commonly used (Ashburner and Friston, 2000; Good et al., 2001; Freedman et al., 2007), central tendency measures may not adequately describe the differences between two clinical groups (Elveback et al., 1970). Although the overlap statistic is referred to as "*d*" in the literature, I use the Greek equivalent "*delta*" (δ) to avoid confusion with "*Cohen's d*". This clashing of terms serves to further illustrate the myriad of statistical approaches available (there are so many they cannot have entirely unique names!)

The complexity of the brain described in chapter 1 and the large number of statistical modelling approaches available (that I have only briefly summarised), led me to conduct the following review. In this review I describe the data, methods, and results from statistical models of brain structure across the lifecourse.

4.2 Methods

I conducted a review of brain structure models concurrently with the reviews of brain MRI databanks (chapter 2) and atlases (chapter 3). A full systematic search of statistical models was not undertaken due to time constraints. Between October 2010 and October 2011, I searched PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and the internet using Google (http://www.google.co.uk/) and Google Scholar (http://scholar.google.co.uk/) with the terms: "Magnetic Resonance Imaging" or "MRI" or "MR" and "brain" and "normal" and "ageing" and "human". I supplemented this search until August 2013 by periodically searching Google with a subset of these terms and by reviewing content alerts distributed by relevant journal articles, e.g. NeuroImage (http://www.journals.elsevier.com/neuroimage/) and Human Brain Mapping (http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1097–0193).

I did not have a formal exclusion process, i.e. I did not keep track of excluded articles, but included all articles I found to model the effects of age on structural brain volumes. That is, whole brain, regional, and voxel volume models across the lifecourse, i.e. I did not include models of diffusion, cortical thickness, or connectivity parameters. As I was interested in assessing brain changes in life, I excluded post–mortem studies. I excluded longitudinal studies as the parameters they assess are inherently different to cross–sectional studies, e.g. within–subject changes from baseline (rather than differences between subjects at different ages) are calculated (Holland et al., 2012). I included baseline cross–sectional analyses from longitudinal studies where available, e.g., Fotenos et al. (2005). Longitudinal data are more robust for defining individual changes in brain structure with time (rather than cross-sectional data that define differences between ages). However, the data to be used in this thesis are cross-sectional (chapter 5); therefore I cannot investigate and/or comment on the best method for longitudinal studies. Given that my interest was in assessing brain structure across the whole lifecourse, I excluded studies that did not

have subjects aged over 60 years. I did not include comparisons between normal and diseased groups, e.g. Alzheimer's diseases, unless there was a specific component that assessed the normal effects of age.

4.3 Results

Seventeen studies were included for review from 591 publications found during the search. The subjects, criteria for normality, image processing, statistical methods, and results from the included studies are described in the following sections.

4.3.1 Subjects

The subjects assessed in the included statistical models of brain structure are described in Table 8.

Ten studies did not report the sampling method used, five reported the use of advertisements, and only two reported that their sample was randomly derived. There was generally an equal representation of gender in all studies. Only three studies had subjects less than 10 years and, in general, most subjects were young to middle aged adults; with the exception of the DeCarli et al. (2005), Jernigan et al. (2001), Peelle et al. (2012), and Ziegler et al. (2011) studies that had larger proportions of subjects aged over 60 years. There were three studies across the majority of the lifespan, Fotenos et al. (2005), Peelle et al. (2012), and Ziegler et al. (2011), that had over one hundred subjects distributed somewhat evenly across ages. All of these subjects were reported to be "normal" and the following sections describe the criteria for normality in each study.

4.3.2 Subject criteria for normality

The criteria for normality provided in each study are described fully in Appendix AI.2, "Criteria for normality in each normal ageing brain volume statistical model". The definition of normality, and the methods used to define it (e.g. questionnaire, physical tests), were different between studies. However, most studies reported that the images of each subject were reviewed by a radiologist or neuroradiologist and

found to contain no overt abnormalities. Following this, each study processed their images as described in the next section.

Table 8. Studies that reported brain volume models across the lifecourse

Author	Sampling method	*n* Gender[1]	Age[2]
Allen et al. (2005)	Advertisements	44 F, 43 M (87)	22–88; 48, 19
Blatter et al. (1995)	Advertisements	105 F, 89 M (194)	16–65; ., .
Courchesne et al. (2000)	Advertisements	37 F, 79 M (116)	1–80; 21, 20
DeCarli et al. (2005)	Random[3]	1133 F, 948 M (2081)	34–96; 62, 10
Fotenos et al. (2005)	.	160 F, 112 M (272)	18–95; 50, 7
Ge et al. (2002)	.	32 F, 22 M (54)	20–86; 47, 19
Giorgio et al. (2010)	.	35 F, 31 M (66)	23–81; 37[c], .
Good et al. (2001)	Advertisements	200 F, 265 M (465)	17–79; 30[c], .
Gur et al. (1991)	.	35 F, 34 M (69)	18–80; 41, 20
Jernigan et al. (2001)	.	41 F, 37 M (78)	30–99; 64, 17
Kruggel (2006)	.	248 F, 254 M (502)	16–70; 30, 10
Li et al. (2012)	.	38F 38 M (76)	19–70; 43, 16
Long et al. (2012)	.	38 F, 39 M (77)	22–89; ., .
Peelle et al. (2012)	.	206 F, 214 M (420)	18–77; ., .
Sowell et al. (2003)	Random[5]	86 F, 90 M (176)	7–87; 30, 22
Walhovd et al. (2005)	Advertisements	40 F, 33 M (73)	20–88; 52, .
Ziegler et al. (2011)	.	305 F, 242 M (547)	19–86; 48, 17

Note: .=I was not able to obtain this information; [1]number of females, males (total); [2]Range; mean, standard deviation (SD); [3]The census of people aged over 20 years in Framingham, Massachusetts, USA was stratified by family size and precinct of residence and then ordered by address; subjects were then systematically drawn from this list (Dawber et al., 1951). [4]Median; [5]Subjects were randomly selected from a telemarketing list of families in the local community.

4.3.3 Image acquisition and processing methods

The image processing methods varied considerably between studies (Table 9). Automated and semi–automated image processing methods have been shown to be in general agreement (Lehmann et al., 2010) and most studies reported some kind of validation or reference to validation. However, because a method has worked well in one group of subjects, it does not mean that it will work well in another group. Each

step of processing (whether automatic or manual) should be thoroughly checked in all new studies, otherwise bias may be introduced (Nordenskjöld et al., 2013).

But perhaps the biggest confounding measure between studies is the choice of registration space and points. Both of these have been shown to significantly affect results (Buckner et al., 2004; Peelle et al., 2012). Different studies have different aims and their own reasons for using one processing method or another but, given the effect these choices have on results, it would be preferable to have a standard protocol to sensibly collate results and reliably define how the brain changes with age. The results from studies that reported models of whole brain tissue volumes across the lifecourse are collated in the next section.

4.3.4 Statistical methods used in brain volume models

The statistical methods used in models of grey matter (Table 10), white matter (Table 11), and cerebrospinal fluid (Table 12) volumes across the lifecourse, and the results from these models, are described here.

From Table 10 there was relative agreement between studies that grey matter volume generally decreases with age. However, the general rate and pattern, i.e. linearly or nonlinearly, at which it decreases did not appear consistent between studies. There was a similar pattern in white matter between studies, it is apparent that white matter generally decreases with age but the rate and pattern was again not apparently consistent (Table 11).

The results in Table 12 are in concordance with a general increase in cerebrospinal fluid with age. Again, some studies report a more dramatic difference between ages than others. The confidence intervals of coefficients were not given in all studies and so tests of heterogeneity, e.g. forest plots and chi-square tests, were not possible for tissue or CSF statistics.

These results (Table 10, Table 11, and Table 12) are all general (mean) differences. Values at other distributional locations, e.g. 5^{th}/ 95^{th} percentile ranks, were only reported in one study. Kruggel (2006) provided the 5^{th} and 95^{th} percentile rank values of normalised brain volumes for males and females aged 16–70 years. These percentile ranks corresponded to the whole age range, i.e. not for each decade or year. The distributions of brain image data may be visualised in "Q-Q" plots and if

the data do not lie on a straight diagonal line they may be transformed (as in Kruggel, 2006) to appear Gaussian.

Table 9. Image acquisition and processing methods in statistical models of brain structure across the lifecourse

Author	Tesla	Sequence	Processing tool	Registration space	Registration points
Allen et al.	1.5	T1	Brainvox, hand tracing	Native AC–PC line	6
Blatter et al.	1.5	T1, T2	ANALYZE	Native	N/A
Courchesne et al.	1.5	T2	SEGMENT (semi–automated)	Native	N/A
DeCarli et al.	1	PD, T2	In–house (semi–automated)	Native, anatomic standard space	"Rigid rotation"
Fotenos et al.	1.5	T1	FSL	Study specific Talairach	12
Ge et al.	1.5	PD, T2	3DVIEWNIX	Native	N/A
Giorgio et al.	1.5	T1	FSL	Study specific MNI152	Nonlinear
Good et al.	2	T1	SPM 99	Study specific MNI152	12 and nonlinear
Gur et al.	1.5	T2	In–house (semi–automated)	Native	N/A
Jernigan et al.	.	T1, T2 FSE	In–house (semi–automated)	Native across sequences	"Resectioned in a standard coronal plane"
Kruggel	3	T1	In–house (automated)	Talairach	12
Li et al.	1.5	T1	FreeSurfer	Native	N/A
Long et al.	3	T1	FreeSurfer	FreeSurfer brain atlas	"aligned"
Peelle et al.	3	T1, T2	SPM	Native, custom MNI152 atlas	Nonlinear
Sowell et al.	1.5	T1	AIR, in–house (semi–automated)	ICBM[1]	12
Walhovd et al.	1.5	T1	FreeSurfer	Aseg Atlas	"Optimal linear and nonlinear transforms"
Ziegler et al.	1.5, 3	T1	SPM	MNI152	12 and nonlinear

Note: SPM=Statistical Parametric Mapping; FSL=(University of Oxford Centre for) Functional MRI of the Brain Software Library; .=I was unable to obtain this information; PD=proton density; [1]Tissue segmentation was conducted in ICBM space but volumes were calculated in subject's native space; FSE=fast spin echo; AC–PC=anterior–posterior commissure.

Table 10. Models of grey matter volume by age: statistical methods and results

Dependent variable	Method (author)	Reported result[1]	
Absolute grey matter volume	Linear regression (Good et al., 2001)	β=–0.0039L (ρ=–0.699),	*P*<0.0001
	Linear regression (Courchesne et al., 2000)	ρ=–0.75 (β=.),	*P*<0.001
	Linear regression (Kruggel, 2006)	β=–1.562mL,	*P*<0.001
	Linear regression (Giorgio et al., 2010)	ρ=–0.84 (β=.),	*P*<0.001
	Nonlinear regression (Sowell et al., 2003)	β=0.60mL*,	*P*=0.01
	Correlation (Ge et al., 2002)	ρ=–0.48,	*P*=0.0004
	Correlation (Li et al., 2012)	ρ=–0.55, (males) ρ=–0.47, (females)	*P*<0.001 *P*=0.003
	t–test (Ge et al., 2002)	717.8mL–626.1mL,	*P*=0.001
Normalised grey matter volume[2]	Linear regression (Good et al., 2001)	β=.,	*P*<0.001
	Linear regression (Kruggel, 2006)	β=–0.008,	*P*<0.001
	Linear regression (Fotenos et al., 2005)	ρ=–0.91 (β=.),	*P*<0.001
	Linear, nonlinear linear regression (Walhovd et al., 2005)	ρ=–0.78, Nonlinear β=.,	*P*=0.000 *P*>0.05
	Correlation (Ge et al., 2002)	ρ=–0.57,	*P*=0.0002
	Spearman's correlation (Jernigan et al., 2001)	ρ=–0.53,	*P*<0.001[C,TSV]
	t–test (Ge et al., 2002)	52.5%–47.6%,	*P*<0.0001

Note: [1]results from regressions are shown, "beta (β), *P*–value"; correlations are shown, "rho (ρ), *P*–value; and t–tests are shown, "mean difference (between young and old adults), *P*–value" where obtainable; .=unobtainable by me; L=litres, mL=millilitres; *=in a nonlinear model this corresponds to a decrease until mid to late adulthood before levelling off; [2]unless otherwise stated volumes were normalised by total intracranial volume; C,TSV=cortical grey matter normalised by total supratentorial volume.

Table 11. Models of white matter volume by age: statistical methods and results

Dependent variable	Method (author)	Reported result[1]	
Absolute white matter volume	Linear regression (Good et al., 2001)	ρ=–0.571 (β=.),	*P*>0.30
	Linear regression (Kruggel, 2006)	β=–0.706 mL,	*P*>0.05
	Linear, nonlinear regression (Courchesne et al., 2000)	ρ=–0.39 (β=.), Nonlinear –β=.,	*P*>0.05 *P*<0.01
	Nonlinear regression (Sowell et al., 2003)	β=–1.94 mL*,	*P*<0.000001
	Correlation (Ge et al., 2002)	ρ=–0.42;	*P*=0.001
	t–test (Ge et al., 2002)	499.9mL–445.3mL,	*P*=0.003
Normalised white matter volume[2]	Linear regression (Kruggel, 2006)	β=–0.0002,	*P*>0.05
	Linear regression (Fotenos et al., 2005)	ρ=–0.25 (β=.),	*P*<0.001
	Linear, nonlinear regression (Walhovd et al., 2005)	ρ=–0.51, Nonlinear β=.,	*P*=0.000 *P*<0.003
	Correlation (Ge et al., 2002)	ρ=–0.35,	*P*=0.009
	Spearman's correlation (Jernigan et al., 2001)	ρ=–0.63,	*P*<0.001[TSV]
	t–test (Ge et al., 2002)	36.6%–34.0%,	*P*<0.02

Note: [1]results from regressions are shown, "beta (β), *P*–value"; correlations are shown, "rho (ρ), *P*–value; and t–tests are shown, "mean difference (between young and old adults), *P*–value" where obtainable; .=unobtainable by me; L=litres, mL=millilitres; *=in a nonlinear model this β corresponds to an increase in childhood that levels off in middle age and decreases after middle age (inverted U shape); [2]unless otherwise stated volumes were normalised by total intracranial volume; TSV=normalised by total supratentorial volume.

Table 12. Models of cerebrospinal fluid volume by age: statistical methods and results

Dependent variable	Method (author)	Reported result[1]	
Absolute cerebrospinal fluid volume	Linear regression (Good et al., 2001)	β=0.0019L (ρ=0.614),	*P*<0.001
	Linear regression (Courchesne et al., 2000)	ρ=0.84 (β=.), (males)	*P*<0.001
		ρ=0.88 (β=.), (females)	*P*<0.001
	Linear regression (Kruggel, 2006)	β=1.54 mL,	*P*<0.001
	Linear, nonlinear regression (Sowell et al., 2003)	β=0.49 mL,	*P*<0.01
		Nonlinear β=0.33 mL,	*P*=0.06
	Correlation (Gur et al., 1991)	ρ=0.74,	*P*<0.0001
Normalised cerebrospinal fluid volume[2]	Linear regression (Good et al., 2001)	β=.,	*P*<0.001
	Linear regression (Courchesne et al., 2000)	ρ=0.81 (β=.),	*P*<0.001
	Linear regression (Kruggel, 2006)	β=0.001,	*P*<0.001
	Linear, nonlinear regression (Walhovd et al., 2005)	ρ=0.69,	$P=0.000^{V}$
		Nonlinear β=.,	$P<0.003^{V}$
	Correlation (Gur et al., 1991)	ρ=0.76,	*P*<0.001
	Spearman's correlation (Jernigan et al., 2001)	ρ=0.83,	$P<0.001^{S,\ TSV}$
	Spearman's correlation (Jernigan et al., 2001)	ρ=0.74,	$P<0.001^{V,\ TSV}$

Note: [1]results from regressions are shown, "beta (β), *P*–value"; correlations are shown, "rho (ρ), *P*–value; and t–tests are shown, "mean difference (between young and old adults), *P*–value" where obtainable; [2]unless otherwise stated volumes were normalised by total intracranial volume; V=ventricle cerebrospinal fluid; S=sulcal cerebrospinal fluid; TSV=normalised by total supratentorial volume.

As well as whole and regional brain volumes, several studies reported voxel–based statistics; these are described in the next section.

4.3.5 Voxel–based statistical methods

The voxel–based statistics reported in these studies are shown in Table 13.

Table 13. Voxel–based statistics reported in brain structure models across the lifespan

Author	Voxel–based statistics
Good et al.	*t*–values, ρ
Giorgio et al.	*t*–values, ρ
Long et al.	*P*–values
Peelle et al.	Means, *t*–values, β/*m*
Sowell et al.	β/*m*, peak ages*
Ziegler et al.	*t*–values, β/*m*, percent reductions

Note: *of nonlinear regression curves.

All of the statistics listed in Table 13 are derived from central tendency measures, e.g. mean/ median. All but ρ ("rho", a measure of correlation) require the parametric assumptions of equal variance and Gaussian distributions (Cohen, 1988).

4.4 Discussion

According to this review, the most commonly applied statistical models in structural brain imaging are from the parametric-frequentist family of methods. That is, I found the overwhelming majority of studies to calculate mean differences in brain structure across the lifecourse. Nonparametric-frequentist methods were reported in one study (Jernigan et al., 2001). All but one of the studies I found reported measures other than central tendencies, i.e. other than means and medians (Kruggel, 2006). Central tendency models are useful for indicating general trends in brain structure across life but if their assumptions, e.g. random samples/ equally distributed data, are not met these may not well describe changes at the extremes of normal ageing. In order to provide useful clinical metrics for diagnoses of pathology, the extremes of brain structure need to be adequately defined (Farrell et al., 2009). While earlier studies could only reliably define central tendencies of brain structure due to the limited amount of data available, there is an increasing amount of brain image data becoming available so that extremes may also be defined (Poline and Poldrack, 2013).

I found one study that attempted to define the extremes of brain structure with statistical modelling (Kruggel, 2006). In that study the limits of normal brain structure were estimated by artificially transforming the data to be distributed approximately

Gaussian and then calculating the mean values minus two SD. By transforming data to the Gaussian distribution the true limits of brain structure may be misrepresented, e.g. increased variance at older ages means that the limits of brain structure are likely greater here than at younger ages (Elveback et al., 1970; Farrell et al., 2009). In this instance, nonparametric methods may be more appropriate for defining the limits of brain structure.

Nonparametric methods have been proposed and widely adopted in functional imaging (Nichols and Holmes, 2002) however they have been implemented far less frequently in structural imaging (Rorden et al., 2007; Yang et al., 2011). This may be because of the perceived decrease in "power" in nonparametric models (Hogg and Tanis, 2010). However, if parametric methods are applied in situations where their assumptions are not met they may provide unreliable results, e.g. if data are not distributed Gaussian their "power" may actually overinflate results (Elveback et al., 1970). I found few studies to report whether their data met the assumptions of the models they used and so, at least according to this review, the robustness of these assumptions and their effects in lifecourse structural brain image data are largely unknown.

Prior to constructing these models, each study defined a set of characteristics by which to classify their subjects as "normal". While the statistical methods for modelling data were generally consistent across studies, methods for determining subject normality differed greatly. Although this is a tricky definition to make, others have noted the adverse effects of differing criteria for normality, e.g. unrepresentative effects of age from "super-healthy" older subjects, and attempted to create standard definitions that may reduce inconsistency of results between studies (Mazziotta et al., 2009). Further to differing criteria for subject normality, there were a range of different image processing tools and methods employed across studies. In conjunction with the varying criteria for normality and robustness of parametric methods (given the underlying data generation and distribution), these differences in image processing methods may explain why this review did not find the effect size of age on brain structure to be consistent across studies.

This review used formal search criteria, however, I did not maintain a record of all excluded studies nor did I perform extensive manual searching throughout the search period. This was therefore not a systematic review so there is a reasonable chance that I did not include all relevant studies. I did check another recent review

(Peelle et al., 2012), and this reported only a couple of additional studies. The majority of cross-sectional studies reported in that review were also found to use the parametric-frequentist methods. In the present review I did not assess longitudinal studies and so, although collated from a number of different sources, these results may not indicate actual changes in brain structure through time but only normal differences between age groups.

These mean differences were defined with the aim of highlighting pathological changes in brain volume (Courchesne et al., 2000). If data are, as expected by parametric methods, distributed equally Gaussian then mean differences will generally be replicated across the spectrum of ageing, e.g. at the clinical limits (Farrell et al., 2009). According to this review it was unclear whether these assumptions were robust across ageing. It was therefore not known if mean-based (parametric) models were applicable at the limits of brain structure.

I proposed to test these assumptions in statistical models of whole brain, regional, and voxel-wise volumes across the lifecourse. Further, given the previous focus on central tendency measures, I sought to determine whether these apply at the extremes of normal brain structure or whether alternative, percentile rank methods are required. Before I did this I assessed various methods of preparing brain MRI data for subsequent statistical analysis. This assessment is described in the next chapter.

5. Brain structure MRI processing across the lifecourse

5.1 Introduction

Given the age-related changes in brain structure that I described in previous chapters, specific MRI processing methods are required at different stages of the lifecourse. But regardless of the age of a subject and even within a single sequence, MR images contain a large amount of data about the brain. For example, T1–weighted images contain whole brain grey matter, white matter, and cerebrospinal fluid volumes, as well as ventricular, cerebellar, lobular, and regional, e.g. hippocampus, volumes. However, these data are not available immediately after scanning; several mathematical operations are required even before the widely recognisable MR image in Figure 1 (page 4) is formed.

The signal detected by a MRI scanner is termed "*free induction decay*" (FID) and this contains information about frequency and phase, i.e. the strength of a magnetisation signal over time. After receiving a MR "pulse", different tissues emit different signal strengths and lose magnetisation at different rates and it these properties that generate the contrast seen in images such as Figure 1. The location and intensity (brightness) of a voxel in a MR image are obtained from the FID signal via a pair of "Fourier Transforms". The mathematics of Fourier Transforms are beyond the scope of this thesis however they, and a wide range of MR physics, are discussed in great detail elsewhere (McRobbie et al., 2007). For present purposes, the point is that different structures in the brain, e.g. grey and white matter, have different magnetic properties and these signals can be mathematically transformed to form grey scale (MR intensity) images such as Figure 1.

Once the MR intensity image has been formed there are several further steps required to obtain tissue and substructure volumes of interest. The first of these steps is generally "*brain extraction*", also known as "*skull stripping*" or "*scalping*". This is basically the removal of all structures in an image, e.g. skull, neck muscles, that do not form part of the brain. Several methods have been proposed to do this manually (Woods, 2005), semi–automatically (Shattuck and Leahy, 2002), and fully–automatically (Smith, 2002; Avants et al., 2008).

Modern brain images may have upwards of 218 acquisition slices and manually drawing around the cortical boundary in each of these slices can take

upwards of an hour and a half per subject (1st year of PhD, David Alexander Dickie). This is clearly not an efficient use of time. At the opposite end of the spectrum is fully automatic brain extraction. Although these methods take almost no human time (they take computational time) and have been validated in some cohorts (Smith et al., 2002), there are concerns that they may not always be reliable (Nordenskjöld et al., 2013). I therefore tested several brain extraction methods in the subjects here to get the best balance between accuracy and efficiency.

Brain extraction is employed for various reasons, the most relevant to the present work being for easier/ optimal extraction of the cortex and subregional volumes, tissue classification and spatial normalisation ("registration"). Tissue classification, based on the intensities and locations of voxels (explaining the need for the removal of skull that has similar intensities to white matter), has been fairly successfully automated across adulthood, including in old age (Zhang et al., 2001; Fotenos et al., 2005). This replaces MR intensities with a standard range representing proportion of a particular tissue within each voxel, i.e. 0 for no grey matter, 1 for entirely grey matter. One reason for doing this is to allow sensible comparison of subjects between scanners and acquisition parameters (that, even for the same tissues, lead to different MR intensity ranges). Standardised voxel values mean that subjects acquired from different scanners, or even on the same scanner but with slightly different acquisition parameters, can be sensibly compared. However, given that tissue classification is partially based on MR intensities, there is the potential for inaccuracies in older subjects with white matter hypointensities (hyperintensities on T2 images that are potentially lesions) that are a similar intensity to grey matter. I therefore assessed another image standardisation method that does not classify voxels but match their range with that of an atlas (intensity range standardisation).

The vast amounts of data needed to effectively understand the brain are unfeasible to collect in one scanner/ centre. However, MR signal intensities, unlike CT Hounsfield units, do not have a universal meaning and so it is difficult to combine data from one scanner that may have a grey matter value range of 3000 to 4000 with another scanner that has a range of 500 to 1500. Given the potential issues with tissue classification I assess another approach which is to normalise the entire range of individual subject voxel values into a reference or template range. This method was previously validated to provide more consistent reporting rating between MR brain images from multiple sclerosis (MS) and tumour patients (Nyúl and Udupa, 1999). I

implemented this method to determine whether it could provide a standard scale of values for the subjects here (who came from multiple scanners) while not unduly altering brain structure.

As well as providing a standard intensity scale, atlases may also provide a standardised spatial framework for images. As discussed in chapter 3, a standard space is required for voxel–based image analyses and there are a variety of registration methods for transforming subjects to this space. Three point registration overlays the centres of images; 6 point transformation overlays the centres and aligns the X (left–right), Y (anterior–posterior), Z (inferior–superior) axes of images; 9 point transformation overlays the centres, aligns the XYZ axes, and rescales images along the XYZ axes; and 12 point transformation overlays the centres, aligns the XYZ axes, rescales the XYZ axes, and rescales the central axes between XYZ. These transformations maintain within brain structure variance, e.g. lateral ventricle volumes, but may not always account for head size differences that occur in axes other than XYZ and the three axes between XYZ, e.g. the axis at X-10 degrees/ Y+90 degrees.

Nonlinear registration is based on directly altering the location of individual voxels rather than via alterations based on their parent axes (Avants et al., 2008). This means that a much higher degree of similarity between a subject and atlas may be achieved, e.g. head size differences are fully accounted for. However, within brain variance of interest, e.g. lateral ventricle volumes, are also warped to be almost the same as the atlas. As discussed in chapter 1, this may be appropriate for functional imaging studies that seek to accurately overlay specific structures, e.g. the amygdala, and compare their functional response. Further, it is also useful for "atlas based segmentation". That is, a manually labelled atlas may be nonlinearly registered to a subject so that these labels can be used to segment the corresponding structures in the subject. This is the method I used to segment regional brain volumes in the subjects studied here.

Nonlinear registration may not be suitable for transforming individual subjects to an atlas in a structural brain image study because it may remove much of the variance in brain structure that the study seeks to quantify, e.g. tissue atrophy. I therefore assessed existing registration methods. Further, I developed my own hybrid linear/nonlinear method in an attempt to mitigate the limitations and exploit the benefits of each method.

The image processing methods discussed were assessed in a range of subjects aged from 0 to 90 years. Data from these subjects were used throughout the remaining chapters and are described in the following sections.

5.2 Subjects

During this work I acquired a number of subjects from across the lifecourse. I did not recruit these subjects myself but acquired them from colleagues within The University of Edinburgh or from publicly available repositories. Starting with infants (0 years) and through to aged adults (60–90 years), these subjects are described in the following sections.

5.2.1 Infant subjects

The infant study was approved by NHS Lothian and National Research Ethics Service, South East Scotland Research Ethics Committee 02 reference: 11/SS/0061 (PI: Dr James Boardman, The University of Edinburgh / MRC Centre for Reproductive Health). Forty–three preterm infant subjects were recruited from the Royal Infirmary of Edinburgh. They were delivered before 36 weeks of gestation and had a birth weight of <1500 grams. Twelve subjects had considerable motion artefact (it is difficult to keep babies still in the scanner) and so were excluded from subsequent image processing and analysis. This meant that I analysed 31 preterm infant subjects. At the time of writing, recruitment was not complete and term equivalent control subjects (delivered after 36 weeks of gestation) were not available. Once recruited, T1–weighted magnetization prepared rapid gradient–echo (MP–RAGE) brain MR were acquired in the sagittal place at 3 T and 1x1x1mm resolution.

To illustrate my intended analysis in the preterm subjects (chapter 9), I used the Montreal Neurological Institute (MNI) 36–44 weeks postmenstrual age normal infant atlas (Fonov et al., 2009) as an intermediary reference. This atlas is a mean image of 36 normal infant subjects; for future analyses I will create a percentile rank atlas of the normal subjects recruited in Edinburgh (target ~15 subjects). Recruitment of the normal young to middle adult sample was complete and these subjects are described in the next section.

5.2.2 Young to middle adult subjects

Eighty normal subjects (40 males, 40 females) aged 25–64 (median 43, IQR 17) years were recruited from advertisements in the Western General Hospital and Royal Infirmary, Edinburgh. National Institutes of Health (NIH) grant R01 EB004155-03 (PI: Dr. Mark E. Bastin, Neuroimaging Sciences, The University of Edinburgh). All subjects gave written informed consent to be assessed and were determined to be normal via medical histories and a battery of cognitive tests. Their ages are detailed in Table 14.

Table 14. Ages of young to middle adult sample

Age group in years	Number
25–34	21
35–44	23
45–54	24
55–64	12
25–64	80

Each subject completed a battery of cognitive tests that were taken from the Wechsler Adult Intelligence Scale III (WAIS-III; Wechsler, 1997a), Wechsler Memory Scale III (WMS-III; Wechsler, 1997b), National Adult Reading Test (NART; Nelson and Willison, 1991), verbal fluency (Lezak et al., 2004), and simple and four–choice reaction time (Deary et al., 2001) tasks. The cognitive domains tested in these tasks are listed in section 7.2.2. The mean NART score was 38 ± 5.6. After determining that they were cognitively normal, brain MRI was acquired from these subjects.

Brain MRI data were acquired with a GE Signa Horizon HDxt 1.5T clinical scanner (General Electric, Milwaukee, WI, USA). The imaging protocol consisted of: axial T2–, T2*–, and FLAIR–weighted sequences, a coronal T1–weighted volume sequence, an axial T1–weighted fast–spoiled gradient echo (FSPGR) sequence, a magnetization transfer (MT–MRI) pulse sequence, and a diffusion MRI protocol, all acquired in Digital Imaging and Communications in Medicine (DICOM). I converted the T1–weighted volume DICOM to NIFTI–1 (http://nifti.nimh.nih.gov /nifti–1) using MRIcron (Rorden, 2010).

5.2.3 Aged adult subjects

Brain MRI from 236 normal subjects and 224 subjects diagnosed with AD (all aged between 55 and 90 years) were acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI; adni.loni.ucla.edu) and the Open Access Series of Imaging Studies (OASIS; http://www.oasis–brains.org/). The normal subjects did not have dementia, but potentially had non–debilitating conditions common in ageing, e.g. hypertension. At the time of the present study, ADNI and OASIS were the only public sources of structural MRI brain scans with clinical metadata that fully represented normal older people (≥60 years) (Dickie et al., 2012). Their demographics are described in Table 15.

Both ADNI and OASIS provided 1.5 T MP–RAGE T1–weighted MR brain images that were acquired in the sagittal plane at approximately 1x1x1mm resolution. Both studies also acquired a range of other MRI sequences, e.g. T2–weighted. However, while theses sequences were available to me from ADNI, they were not available to me from OASIS. OASIS data were acquired in one site whereas the ADNI data were acquired from 60 sites. The full image acquisition parameters are described in Marcus et al. (2007c – OASIS) and Jack Jr et al. (2008 – ADNI).

5.3 Image processing methods

I performed various processing steps for each subject group, some of which were specific and others were applied to all groups. These processing steps are described in the following sections. Brain extraction, which I performed in all subject groups but with varying methods, is described first.

5.3.1 Brain extraction

Brain extraction, sometimes referred to as "skull–stripping", is often the first step in obtaining quantitative data from MR brain images. The complex and variable shape of the human brain means that this task is neither trivial nor amenable to unsupervised, mass computer processing. However, the need for mass amounts of data to effectively understand the brain (Koslow, 2000; Toga, 2002) has led several groups to attempt to partially or fully automate brain extraction (Shattuck and Leahy, 2002; Smith, 2002;

Avants et al., 2008). Since I sought to obtain such a large amount of data, I tested whether these methods were suitable, or could be developed, for the broad subject population studied here (which included infant, aged adult, and young to middle aged adult subjects).

Table 15. Demographics of the ADNI and OASIS subjects

Sample	Age in years	Number of M:F (total)	
ADNI			
Controls	<70	0:0	(0)
	70–74	35:28	(63)
	75–79	25:22	(47)
	80–84	9:8	(17)
	85–89	8:2	(10)
	≥90	0:1	(1)
	Overall	77:61	(138)
AD	<70	16:12	(28)
	70–74	14:17	(31)
	75–79	12:14	(26)
	80–84	16:8	(24)
	85–89	8:7	(15)
	≥90	0:0	(0)
	Overall	66:58	(124)
OASIS			
Controls	<70	7:18	(25)
	70–74	7:19	(26)
	75–79	3:6	(9)
	80–84	4:13	(17)
	85–89	4:9	(13)
	≥90	1:7	(8)
	Overall	26:72	(98)
AD	<70	6:9	(15)
	70–74	10:15	(25)
	75–79	10:13	(23)
	80–84	10:15	(25)
	85–89	3:4	(7)
	≥90	2:3	(5)
	Overall	41:59	(100)

Note: M=males; F=females; AD=Alzheimer's disease

The Brain Extraction Tool (BET; Smith, 2002) is one of the most commonly used fully automated skull-stripping method. BrainSuite is a semi-automated skull-stripping method. Atlas based segmentation is also regularly used and may be implemented with Advanced Normalization Tools (ANTS) registration (Avants et al., 2008). Registration is discussed in detail in section 5.3.5, "Spatial normalisation (registration)".

Since there is less variability among young to middle aged subjects than at the extremes of life, I considered that the ANTS method would perform very well in these subjects. I selected the Montreal Neurological Institute (MNI) 152 brain mask atlas as the reference to register to all subjects. When I acquired this atlas I noted that it was provided in resolutions of 0.5, 1.0, and 2.0mm^3. Although the subject images that I studied here were in approximately 1.0mm^3, I speculated that by downsampling these images to the 2.0mm^3 atlas, I could reduce the manual editing time by half. Notably, this would not unduly alter subjects as I intended to resample the masks back to 1.0mm^3 resolution and apply these masks to the original images. Following brain extraction, regional structures were extracted using atlas-based segmentation.

5.3.2 Regional brain volume extraction: atlas based segmentation

The SRI24–TZO atlas (Rohlfing et al., 2010) consists of a large number of manually delineated subregional brain volumes, e.g. lateral ventricles, supplementary motor area. There are over 100 gyral and subcortical brain regions in this atlas but, given their associations with ageing and disease (Jack Jr et al., 1997; Job et al., 2002), I only used the subcortical grey matter regions in my analysis. That is, the left and right hippocampus, amygdala, parahippocampal gyrus, caudate, putamen, and thalamus. Atlas–based segmentation is the process of registering (reshaping) these labels to match the anatomy in new subjects and using the transferred labels to extract regional brain volumes from these subjects. It is a relatively simple process where the structural (T1–weighted) atlas image is registered to the subject's T1–weighted image. This produces a "transformation matrix" (the series of calculations required to reshape the atlas into the shape of the subject) that may be applied to any image, i.e. label image, in the atlas space so that it is then in the subject space (Figure 7).
I used ANTS to reshape (nonlinearly register) the SRI24 structural and label images to each subject. Registration methods are described in detail in section 5.3.5. For now, it

is important to note that the subject was not registered (deformed) to the atlas, but that the atlas was deformed to the subject. If subjects were deformed to the atlas then they would all have volumes biased towards those of the atlas (and not reflective of their own actual volumes). Volumes are calculated by summing the proportions of grey matter in each voxel in the region. Proportions of grey matter are calculated with tissue classification methods; I describe these in the next section.

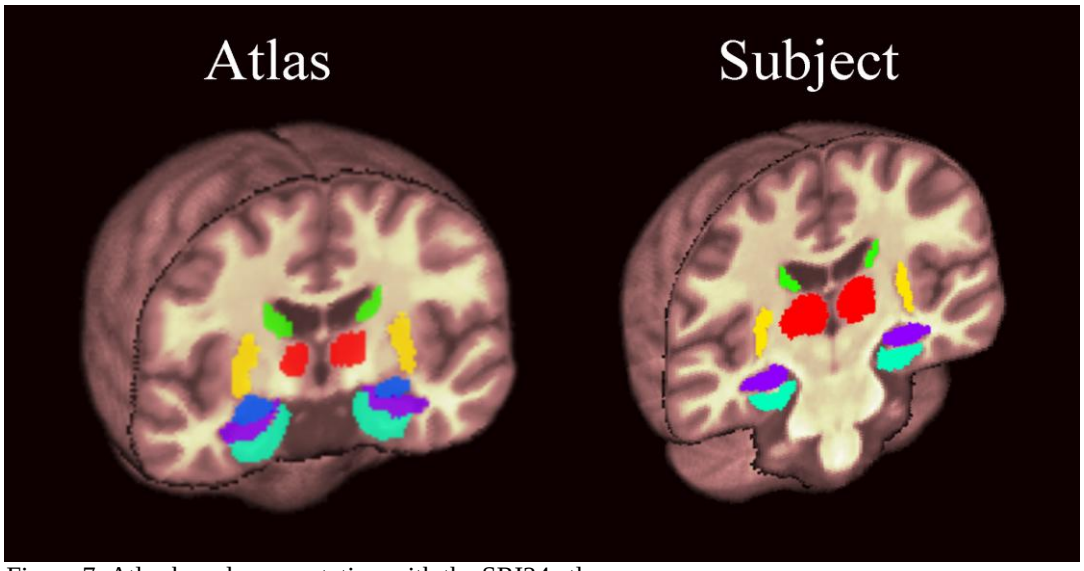


Figure 7. Atlas based segmentation with the SRI24 atlas
The structural atlas image (coloured pink on the left) is nonlinearly registered to the subject's structural image (right) and the resulting transformation matrix is the applied to the coloured labels so that they are also in the shape of the subject's anatomy. In this example the subjects anatomy was such that all regional volumes did not appear on one slice (the amygdala was calculated, it just did not extend to this slice for this particular subject). Green=caudate; red=thalamus; yellow=putamen; royal blue=amygdala; purple=hippocampus; light blue=parahippocampal gyrus.

5.3.3 Tissue classification

Tissue classification and bias correction was performed with the FMRIB's Automated Segmentation Tool (FAST) in a number of previous studies (Zhang et al., 2001; Buckner et al., 2004; Fotenos et al., 2005; Marcus et al., 2007c). Since I aimed to evaluate the statistical methods used in previous volumetric brain studies, I choose to use FAST for tissue classification (limiting the potential for my results being due to differences in image processing). Since all tissue classification methods I could find were based on similar principles and produce similar results (Avants et al., 2008), I

did not conduct an assessment of methods. Rather, I attempted to implement another method for standardising data between scanners.

5.3.4 MR intensity scale standardisation

To normalise the entire range of individual subject voxel values into a reference or template range, the standardised intensity scale (SIS), equation 2, is performed in each voxel

$$\overset{\rho}{V}_{i,j} = \begin{cases} \mu_s + (V_{i,j} - \mu_j)\dfrac{s_1 - \mu_s}{p_{1j} - \mu_j}, \text{ if } m_{1i} \le V_{i,j} \le \mu_j \\ \mu_s + (V_{i,j} - \mu_j)\dfrac{s_{99} - \mu_s}{p_{99\,j} - \mu_j}, \text{ if } \mu_j < V_{i,j} \le m_{2i} \end{cases} \quad (2)$$

where $\overset{\rho}{V}_{i,j}$ is the transformed value of voxel i in subject j, μ_s is the median of the template, $V_{i,j}$ is the original value of voxel i in subject j, μ_j is the median of subject j, s_1 is the 1st percentile of the template, p_{1j} is the 1st percentile of subject j, s_{99} is the 99th percentile of the template, p_{99j} is the 99th percentile of subject j, m_{1i} is the minimum value of subject j, and m_{2i} is the maximum value of subject j. This is in effect shifting the subject's histogram into the scale of the template histogram (Nyúl and Udupa, 1999). I implemented this algorithm in the Matrix Laboratory (MATLAB) computing language.

5.3.5 Spatial normalisation (registration)

My work on spatial normalisation (registration) began with a project in collaboration with an industrial partner, Toshiba Medical Visualisation Systems Europe (TMVSE). During this project I worked on site to assess their registration algorithms in comparison with the freely available and widely used FMRIB's Linear Image Registration Tool (FLIRT) and ANTS algorithms. The TMVSE data are confidential so cannot be reported here but I do describe how the FSL and ANTS algorithms affect brain structure.

Forty–nine randomly selected normal ageing subjects (aged ≥60 years) were taken from the OASIS databank (Marcus et al., 2007c). These 49 subjects had clinical characteristics associated with normal ageing, e.g. hypertension and diabetes, but not dementia. Their mean age was 75.92 years, standard deviation (SD) 8.98 years (range

60–94 years). Approximately 73% of the normal ageing subjects in OASIS were female.

5.3.5.1 Standard space (atlas) for registration

A random selection of half (49) of the normal OASIS subjects was used to create the template standard space (Figure 8):

1. The brain extracted image of each randomly selected subject was linearly orientated (not stretched) to the MNI Colin 27 brain extracted template, i.e. a 6 point linear transformation
2. The mean voxel–wise intensity was calculated in this space
3. Each subject was then oriented and linearly deformed (12 point transform) to this average brain
4. The mean voxel–wise intensity was calculated in this space
5. Each subject was then nonlinearly deformed (Avants et al., 2008) to this average brain
6. The mean voxel–wise intensity was calculated in this space and this defined the template standard space (Figure 8)

3D renderings of the aged standard space and the MNI atlas, both created with MRIcron (Rorden, 2010), are shown in Figure 9. This illustrates that the larger ventricles, sulcal spaces, and overall reduced brain tissue volume associated with ageing (Farrell et al., 2009) is not represented in atlases from younger subjects, e.g. MNI.

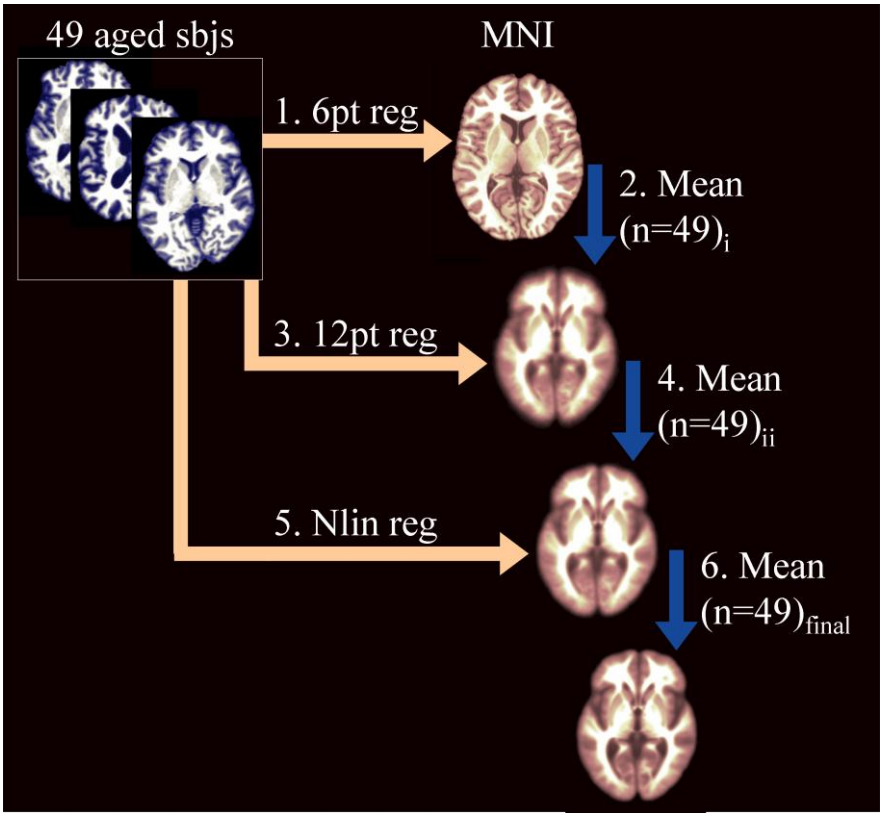


Figure 8. Procedure to create a standard space representative of aged (≥60 years) subjects

This is in the orientation of the commonly used Montreal Neurological Institute (MNI) atlas but was derived from half (49) of the normal OASIS subjects; sbjs=subjects; pt=point; reg=registration; Nlin=nonlinear.

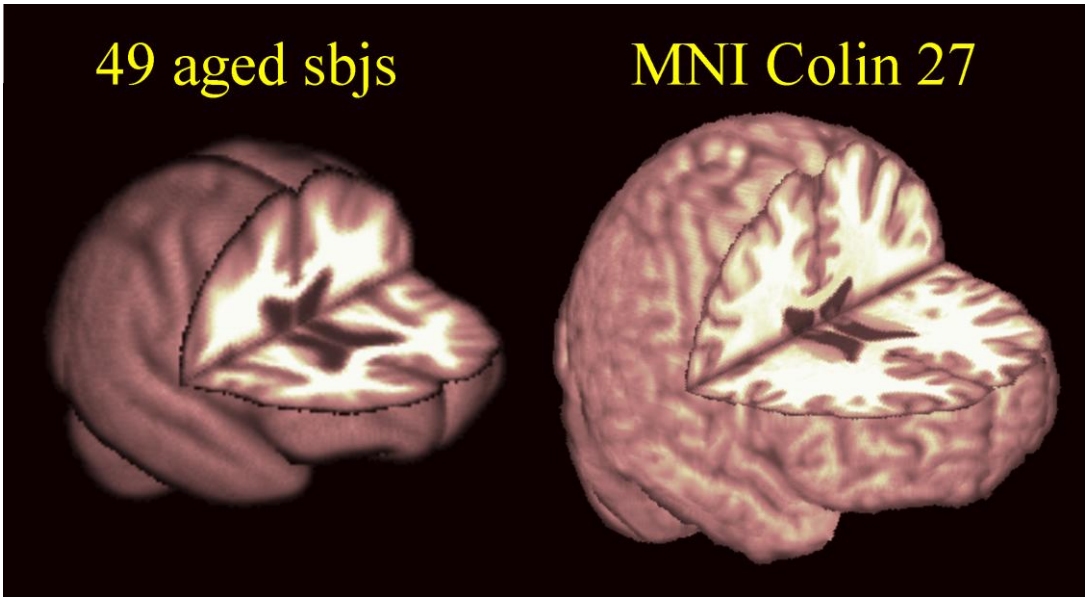


Figure 9. The mean atlas image from 49 aged normal subjects (60–90 years) in the space of the Montreal Neurological Institute (MNI) Colin 27 atlas

The aged atlas is markedly smaller and has larger ventricles and sulcal spaces than the Colin 27 atlas that is based on a ~35 year old subject. The blurriness of the aged atlas is due to it being the mean of 49 different subjects; the MNI Colin 27 atlas is the mean of the same subject scanned 27 times; sbjs=subjects.

I registered the aged subjects to aged standard space (Figure 9) by the methods described in the following sections.

5.3.5.2 Linear registration

I performed linear registration using FLIRT. FLIRT provides a number of linear registration methods with 3, 6, 9, and 12 point transformations (Jenkinson and Smith, 2001; Jenkinson et al., 2002). I registered each subject into standard space with FLIRT using 6, 9, and 12 point transformations.

5.3.5.3 Nonlinear registration

I used ANTS with step–size 0.25 and 100×100×100×25 iterations to perform nonlinear registration. These parameters, specific to the ANTS program, were chosen to optimise similarity between the atlas and subject (Avants et al., 2008).

5.3.5.4 Comparison of registration methods

To compare the results of each registration method I calculated mean and standard deviation images of all subjects after each registration. I then used these to compute the coefficient of variance (CV) in each voxel. CV is calculated by equation 3

$$CV_i = \frac{\sigma_i}{\mu_i} \tag{3}$$

where CV_i is the coefficient of variance in voxel i, σ_i is the standard deviation in voxel i, and μ_i is the mean in voxel i.
I then computed median change in CV after each registration, relative to the 6 pt (lowest) degree of registration.

5.3.6 A novel registration method: “nonlinear surface” registration

Linear registration maintains within brain variance of interest, e.g. ventricle size, between subjects but may not always adequately account for head size differences. Nonlinear registration generally accounts for head size differences but may remove within brain variance of interest, e.g. ventricle size, between subjects. Therefore, I

devised “nonlinear surface” registration by nonlinearly registering each subject’s brain mask (binary image) with ANTS to the standard space mask. The transformations calculated with the masks (Figure 10) were applied to the original MR intensity image of each subject.

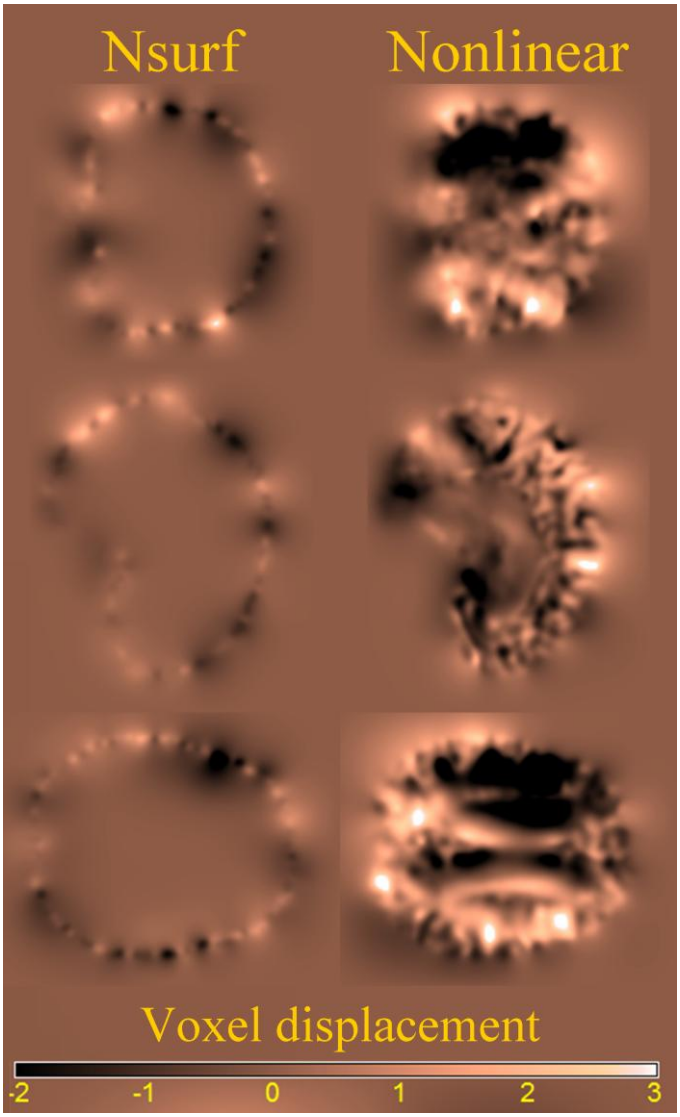


Figure 10. Warps from nonlinear and "nonlinear surface" (Nsurf) registration
In nonlinear registration the entire brain is warped to approximate the atlas brain, in nonlinear surface registration only the overall size and shape of the brain is warped to the atlas (the ratio of within brain structure, e.g. lateral ventricles, is not warped).

5.3.7 Cortical thickness

I used BrainSuite (Shattuck and Leahy, 2002) to extract the cortex from a subset of the aged subjects. BrainSuite extracts the cortex by defining the grey/white (inner) cortical boundary and the grey/ cerebrospinal fluid (outer) cortical boundary (Figure 11).

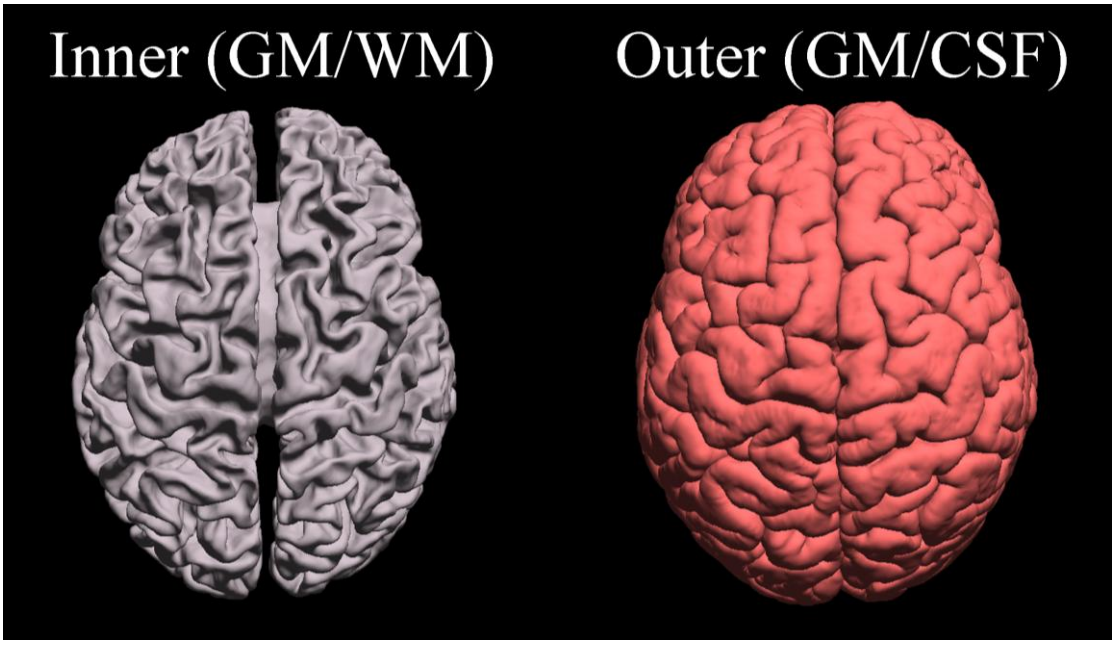


Figure 11. BrainSuite method for measuring cortical thickness
The inner (grey/ white matter) and outer (grey matter/cerebrospinal fluid) cortical boundaries are extracted and the Euclidean distance between them calculated.

Cortical thickness is then measured by calculating the Euclidean distance between the inner and outer boundaries.

5.4 Image processing assessment results

The following sections describe the results from each image processing assessment for each subject group.

5.4.1 Brain extraction

Brain extraction was the first image processing step that I implemented in each subject group, starting with infant subjects.

5.4.1.1 Infant subjects

I first attempted brain extraction in the infant subjects with BET.

5.4.1.1.1 BET

An example of infant subject skull stripping using the default parameter settings in BET is shown in Figure 12. Although the top of the brain was extracted very well, a large amount of non–brain structure remained around and below the optic nerves. By adjusting BET's default settings I was able to improve the brain extraction (lower panel Figure 12). Despite correcting some of the default setting errors, reasonably large amounts of non–brain structure remained, while a very large proportion of brain structure was incorrectly removed. This meant that mass unsupervised skull stripping using BET was not suitable for the infant subjects. Indeed, BET was not designed for use in these subjects. The BET process took less than 30 computer seconds and manual editing took between 30–60 minutes per subject.

5.4.1.1.2 BrainSuite

BrainSuite generally extracted the infant brain very well, but while there were only minor errors in some subjects (Figure 13i), there were more considerable errors in others (Figure 13ii). Further, despite the subjects being from the same scanner and sequence, the high variability in brain structure between subjects meant that the extraction parameters had to be adjusted for each individual. Therefore mass unsupervised skull stripping using BrainSuite was not suitable for the infant subjects.

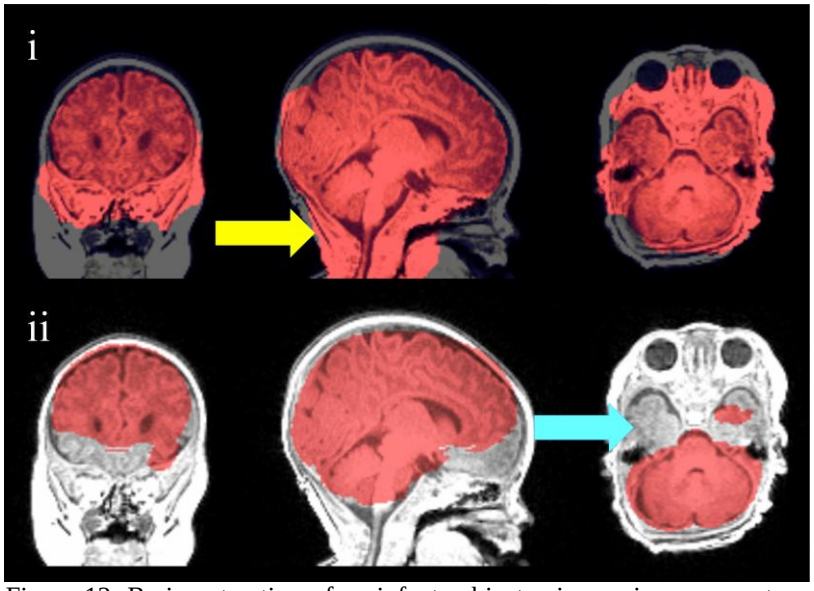


Figure 12. Brain extraction of an infant subject using various parameters in the Brain Extraction Tool (BET)
The extracted brain is highlighted in red and overlaid on the original whole head image. The yellow arrow shows erroneous remaining non-brain tissue and the blue arrow shows erroneous missing brain tissue.

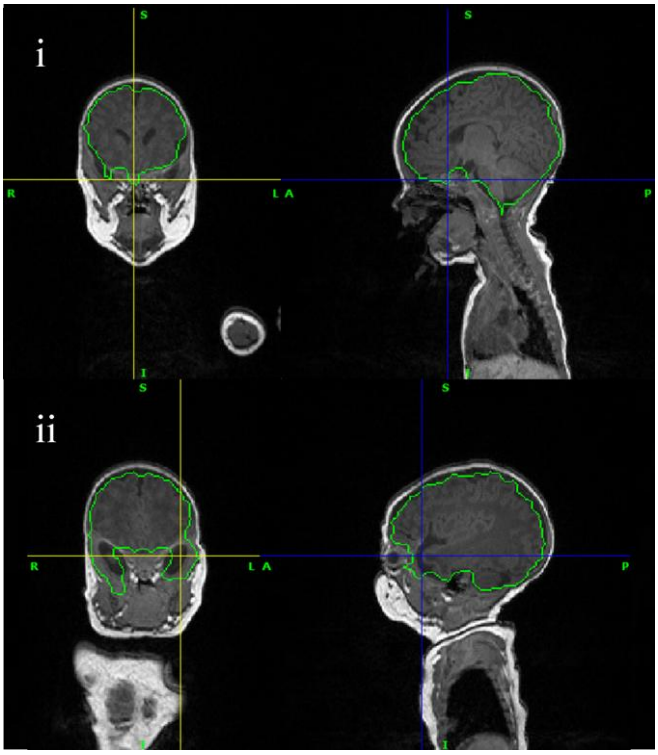


Figure 13. Automatic brain extraction in two separate infant subjects using BrainSuite
The extracted brains are outlined in green and overlaid on the original whole head images. The centres of the crosshairs show where non-brain tissue remained. BrainSuite provided a tool to remove this manually.

The BrainSuite process took less than 30 computer seconds per subject and manual editing took between 20–45 minutes per subject. Figure 14 shows an example BrainSuite result after manual editing.

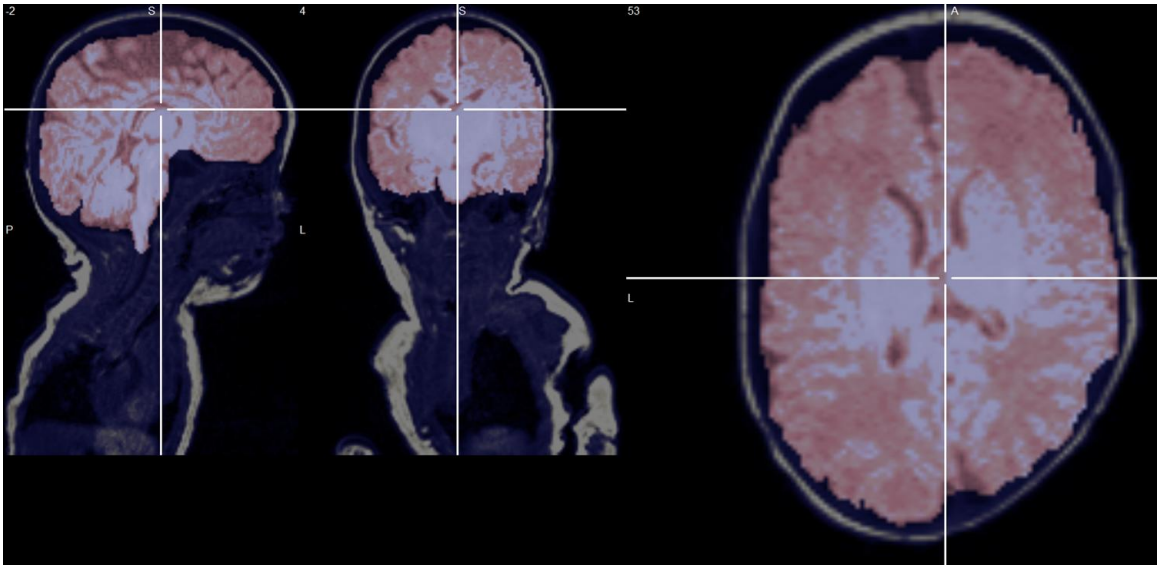


Figure 14. Brain extraction in an infant subject after manual editing in BrainSuite
The extracted brain is highlighted in pink and overlaid on the original whole head image.

5.4.1.1.3 Atlas based brain extraction with ANTS

By diffeomorphically registering the (manually validated) extracted brain from Figure 14 to all other subjects, ANTS performed brain extraction very well. Of the 31 infant subjects, there were two gross errors while the remaining 29 subjects had slight volumes of meninges and skull (Figure 15). The ANTS process took one computer hour and manual editing took between 10–20 minutes per subject. The requirement for manual editing meant that mass unsupervised skull stripping using ANTS was not suitable for the infant subjects.

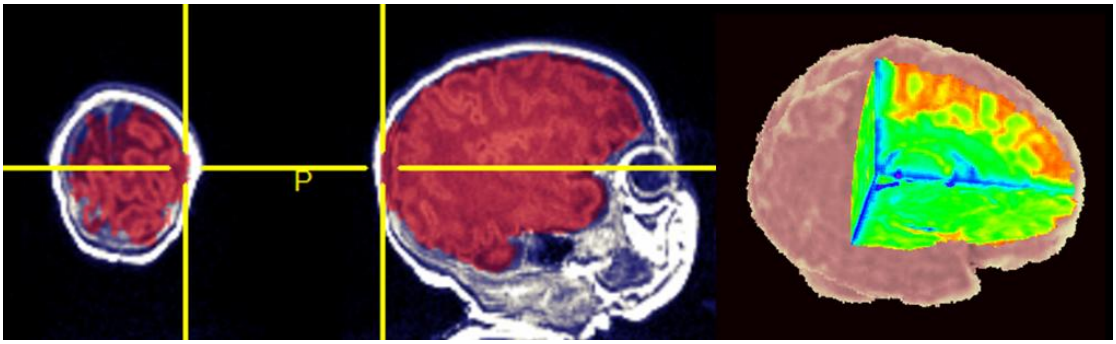


Figure 15. Automatic atlas based brain extraction in an infant subject using ANTS
The extracted brain is highlighted in red and overlaid on the original whole head image. The centre of the yellow crosshairs shows the minor errors that occurred in this method. On the far right is a brain extraction 3D rendering that illustrates the limited errors across the cortex. The green/blue areas are tissue and the red/ yellow areas are cerebrospinal fluid.

Since the high variability between infant subjects is similar to the variability between aged adults, I adopted what I found to be the best method for the infant subjects (atlas based segmentation) for the aged adult subjects.

5.4.1.2 Aged adult subjects

The long computational processing time of ANTS registration (~1 hour) led me to attempt atlas based segmentation in the aged adult subjects with linear registration. Linear registration computes far less transformations (this is discussed in detail in section 5.3.5) and so has a much shorter computational processing time (~2 minutes). However, the reduced number of computations means that differences between the atlas and subject are greater and this adversely effected brain extraction (Figure 16).

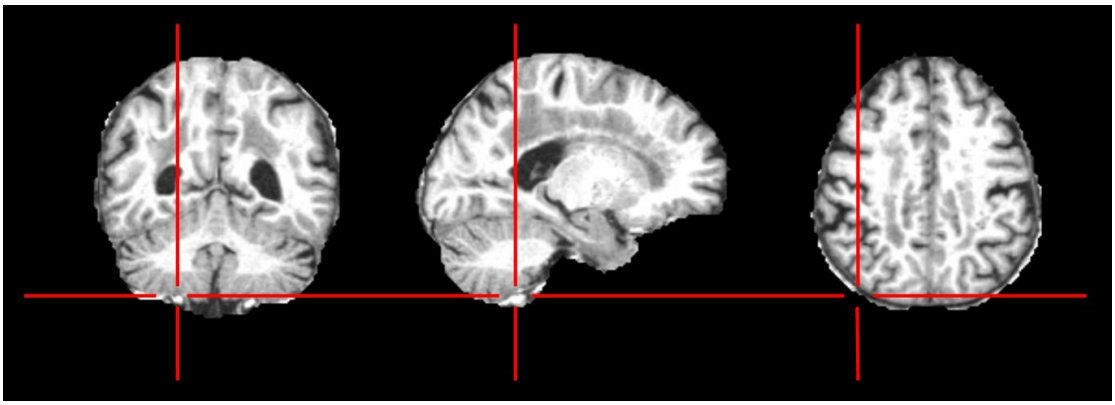


Figure 16. Automatic atlas based brain extraction in an aged subject using 12 point linear registration
This led to several erroneous areas where skull remained, indicated by the centres of the red crosshairs.

These errors and doubled manual editing time (20–40 minutes per subject) determined that atlas based segmentation with ANTS registration was most suitable for the aged adult subjects (Figure 17).

As illustrated in Figure 17, ANTS performed very well for the majority of the 460 aged subjects. This still required approximately 10 minutes of manual editing per subject and, although very infrequent, gross errors were not eliminated entirely (Figure 18).

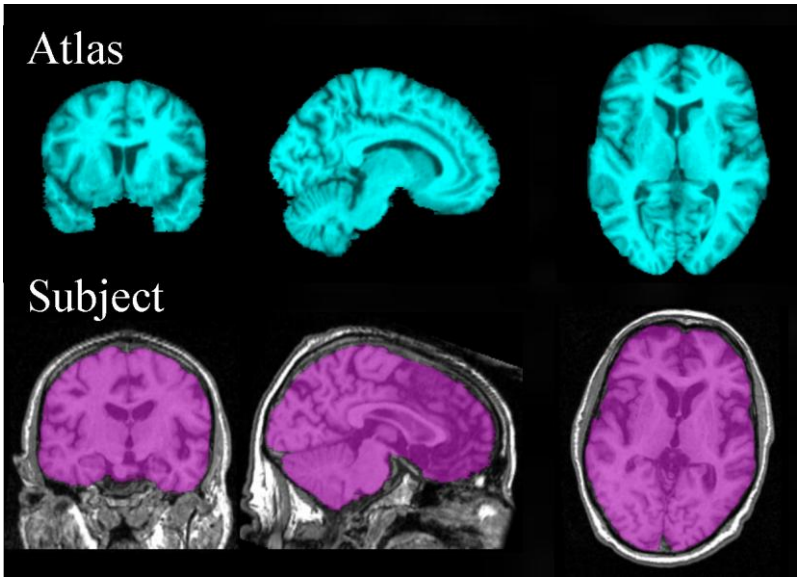


Figure 17. Automatic atlas based brain extraction in an aged subject using ANTS
The atlas in cyan was registered to the subject using ANTS for brain extraction. The extracted brain is highlighted in magenta and overlaid on the original whole head image.

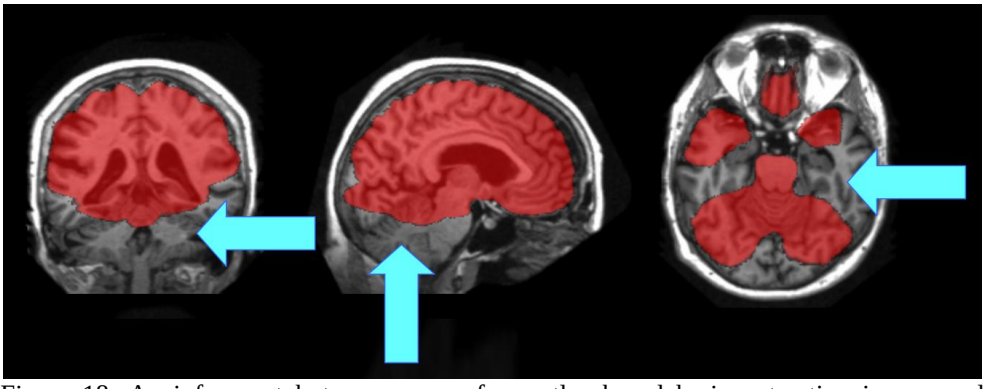


Figure 18. An infrequent but gross error from atlas–based brain extraction in an aged subject using ANTS
The attempted brain extraction is highlighted in red and overlaid onto the original whole head image. The blue arrows show that large amounts of brain tissue have been erroneously missed.

The gross error shown in Figure 18 and the more subtle errors in Figure 17 (at the outer edge of the cortex) meant that mass, unsupervised skull stripping was not suitable for the aged adult subjects. The error rates, computational times, and manual editing times for each brain extraction method are given in Table 16.

Table 16. Error rates and processing times in aged subject brain extraction

Method	Percent of subjects with gross errors	Computational time	Manual editing time
Linear registration	75%	2 minutes	20-40 minutes
ANTS registration	5%	1 hour	10-20 minutes

Note: ANTS=Advanced Normalisation Tools.

The errors in aged adult brain extraction were due in part to the high variance in ageing brain structure but fortunately this problem did not exist across the whole lifecourse. Brain structure was much more homogeneous in younger adult subjects and this allowed me to implement a downsampled method in these younger subjects.

5.4.1.3 Young to middle adult subjects

The downsampled method that I proposed in young to middle adults was generally successful. Manual editing time was reduced by approximately half, to approximately 5 minutes per subject (from 10 minutes in original resolution). However, despite this reduction, my approach did not completely negate the need for manual editing (Figure 19). Therefore, even in these much more homogenous subjects, mass unsupervised skull stripping was not suitable.

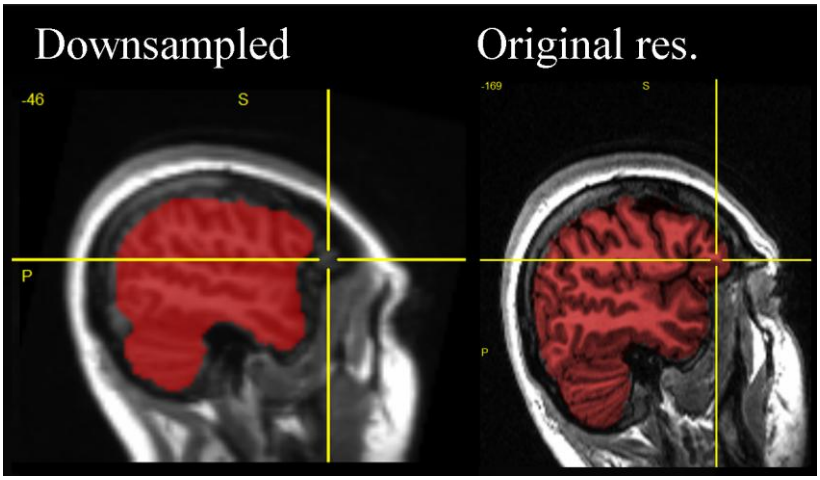


Figure 19. Automatic brain extraction in a middle aged adult using a downsampled atlas
The extracted brain is highlighted in red and on each whole head image. The downsampled mask was edited (see missing brain tissue on the left at the centre of the crosshairs) and then returned to its original resolution to provide the subsequent brain extraction in the original subject (right).

5.4.2 Tissue segmentation

Examples of FAST tissue classification in the aged subjects are shown in Figure 20.

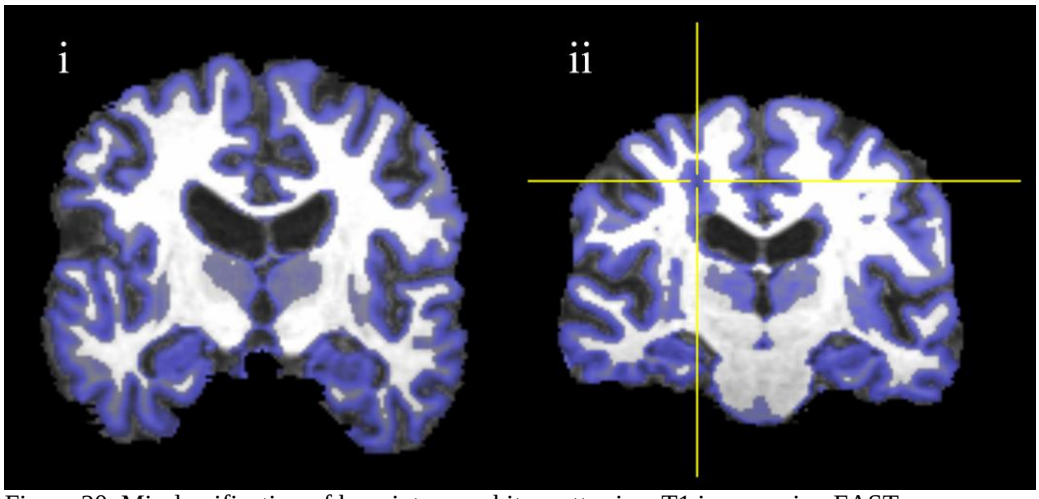


Figure 20. Misclassification of hypointense white matter in a T1 image using FAST
Grey matter has been highlighted in blue showing that subject i was correctly classified, however, this also shows that the classification algorithm has incorrectly classified hypointense white matter as grey matter in subject ii. The misclassifications are highlighted at the intersection of the crosshairs (right panel).

Although FAST differentiated brain tissue from cerebrospinal fluid very well, it sometimes misclassified white matter hypointensities (hyperintensities on T2–

weighted images) as grey matter (Figure 20). Gross errors occurred in approximately 5% of aged adult subjects and were not focused in any one age group but equally spread across the age range. These errors may be corrected using multimodal MR imaging, e.g. the combination of T1-weighted and FLAIR images differentiates white matter abnormalities from GM (Admiraal-Behloul et al., 2005). The use of multimodal MRI sequences and parallel molecular imaging has been shown to provide improved characterisations of brain structure (Toga et al., 2006). However, only T1-weighted images were available for many of the aged adult subjects used here (Marcus et al., 2007c). There were almost no hypointense white matter regions in the young to middle adult subjects. Because tissue intensities are inverted in early life, I did not use tissue classification in infant subjects.

In an attempt to mitigate errors in standardising voxel values while still being able to sensibly combine subjects from different acquisition parameters and scanners, I implemented a MR signal normalisation algorithm (Nyúl and Udupa, 1999).

5.4.3 Standardising the MR intensity scale

To illustrate the standardisation procedure and its effect on brain structure, the original voxel intensities of 25 normal aged subjects (different individuals) from 3 different scanners in ADNI and those of the template are shown in Figure 22. The standardised intensities in Figure 23i show that this method successfully normalised the intensity scales from these three scanners. Figure 23ii shows that this was achieved without unduly altering original brain structure: median, IQR error in grey matter volume following intensity normalisation was 0.011, 0.019 (1.1, 1.9%). Only 25 subjects are shown in Figure 22 to avoid clutter, the tissue segmentation was carried out in 49 subjects (Figure 23ii).

The minimal error in tissue segmentation before and after SIS suggested that the algorithm removed only between scanner variance and not between subject variance (right panel Figure 23). However, as shown in Figure 21, this original algorithm did affect the appearance of brain structure, i.e. it increased tissue contrast. Although this may be aesthetically pleasing, it is an artificial effect that does not reflect the true brain structure (tissue contrast naturally decreases in ageing; Pakkenberg et al., 2003).

Figure 21. Tissue contrast in an aged subject before and after the original standardised intensity scale (SIS) algorithm

After investigating this effect, I found that it was introduced by removing image background values before the standardisation procedure. I therefore updated the algorithm to include image background values during standardisation and only remove these for graphically visualising output (e.g. Figure 24). Tested in 460 different subjects (N_{AD}=224, $N_{Control}$=236) from 60 different sites used in ADNI and OASIS, the updated SIS algorithm standardised voxel values whilst not affecting the original appearance of brain structure (Figure 24).

This validation is not absolute and applies only to the subjects that I assessed. I recommend that, even if already validated in one group of subjects, thorough visual and quantitative testing is done prior to performing subsequent analyses after SIS.

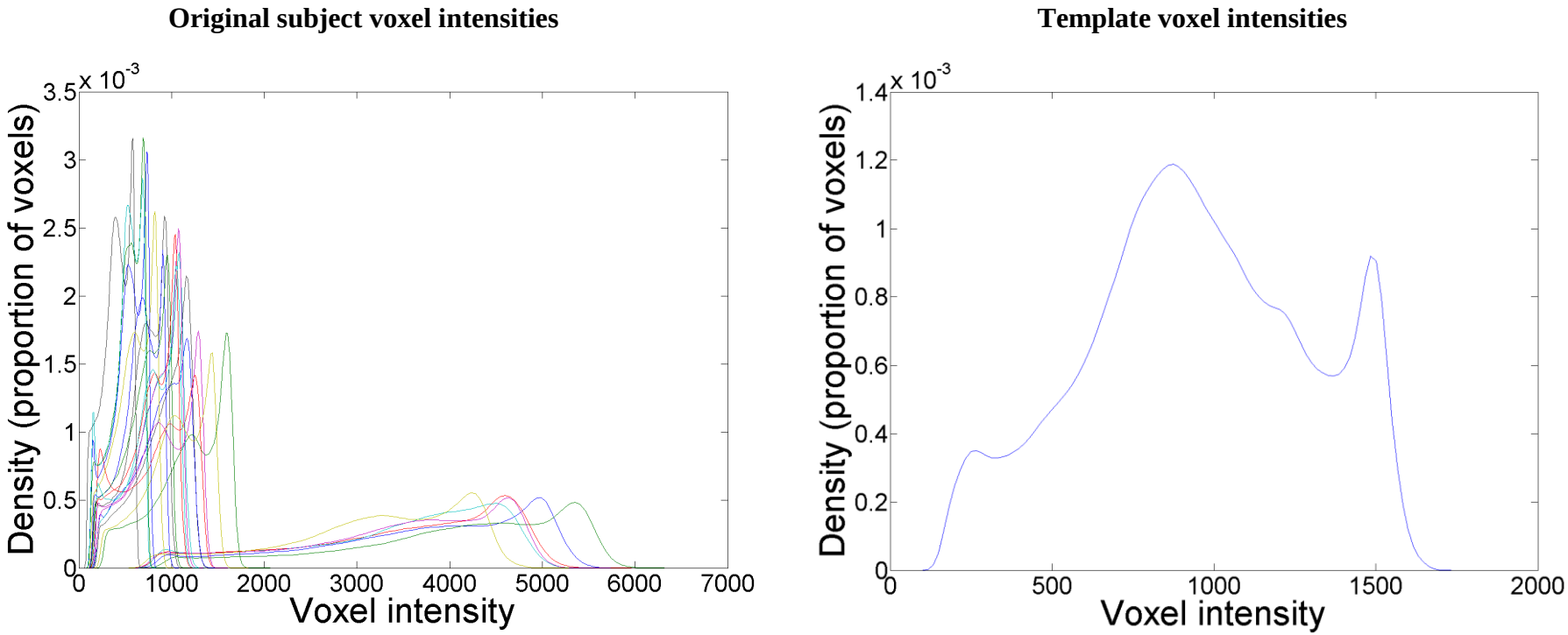


Figure 22. Original voxel intensities of 25 subjects from 3 different scanners and the template voxel intensities
This shows the original (and differing) voxel ranges of each subject (left panel) and the target/ template range (right panel).

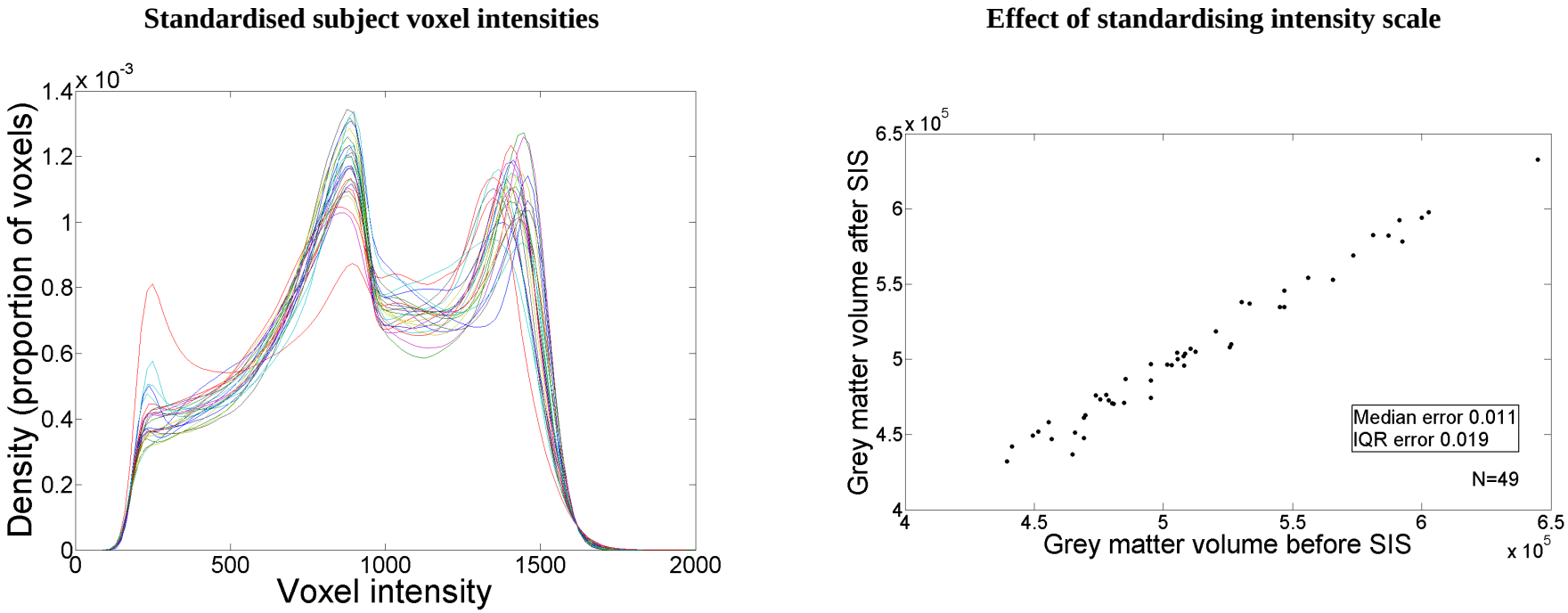


Figure 23. i. Standardised voxel intensities of 25 subjects from 3 different scanners and ii. The effect of standardisation on original brain structure
This shows the standardised voxel ranges of the previously differing subjects (right panel) and the almost null effect this standardisation had on underlying brain structure.

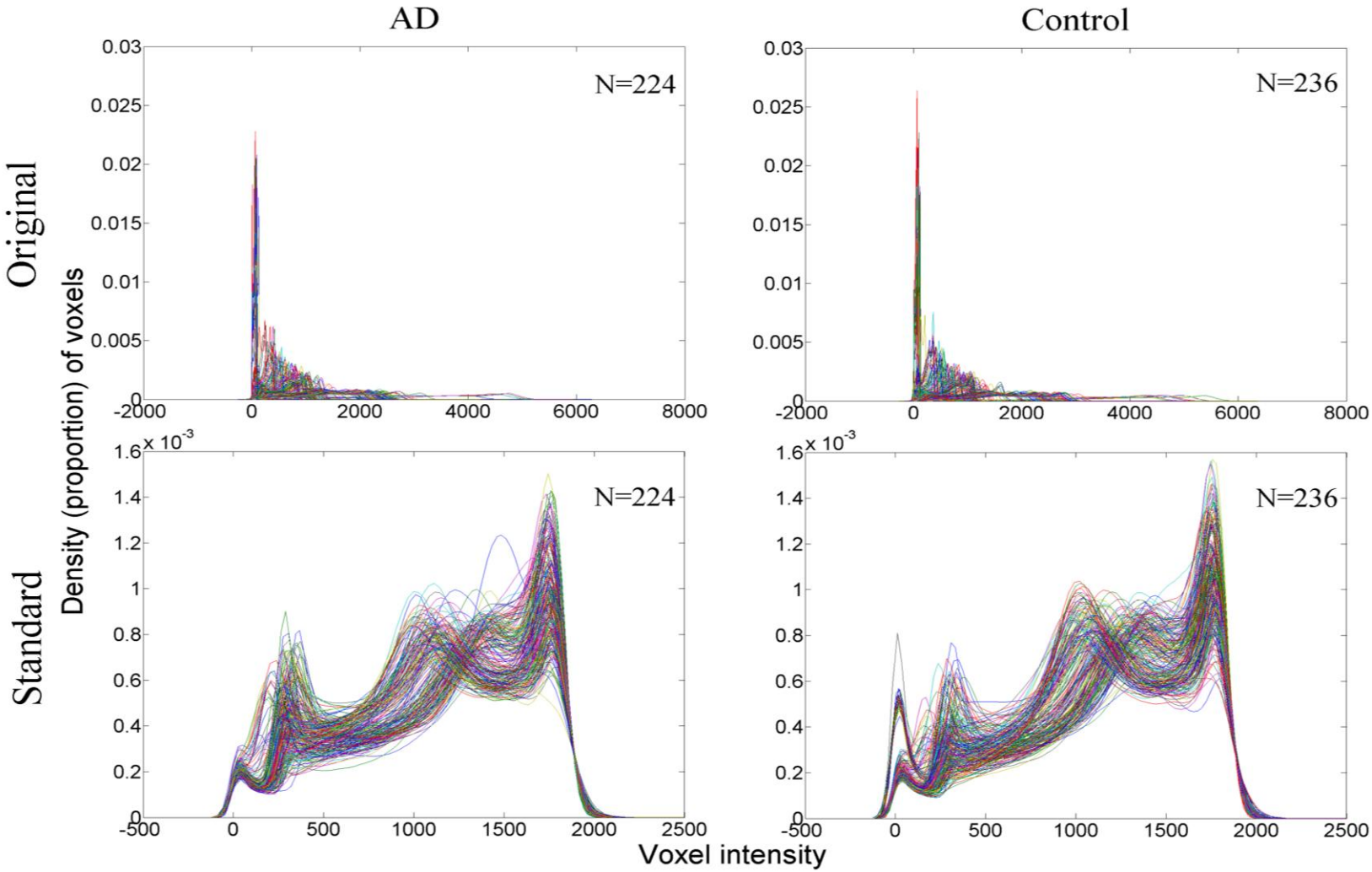


Figure 24. Output from updated standardised intensity scale (SIS) algorithm
The updated algorithm removed background voxels after standardisation, only for graphically visualising output, i.e. these histograms.

5.4.4 Spatial normalisation (registration)

The results from each type of registration are illustrated in Figure 25. Linear registration methods with a lower number of transformations (6 point, 9 point, and 12 point) maintained original anatomical variation but did not always adequately account for head size differences. Nonlinear registration adequately accounted for head size differences however removed almost all of the within brain variance of interest, e.g. lateral ventricle volume. Nonlinear surface registration addressed both of these issues by adequately accounting for head size while still maintaining within brain variance of interest.

CV was 0.45 6pt registration, 0.37 in 9pt registration, 0.36 in 12pt registration, 0.33 in nonlinear surface registration, and 0.25 in nonlinear registration. This meant 9 point (pt) registration reduced CV by 17.7%, 12 pt registration by 20%, nonlinear surface registration by 27.5%, and nonlinear registration by 44.9% (with respect to 6 pt CV). Higher reductions in CV indicate that the registration method is removing more of the natural variance in brain structure between subjects, i.e., nonlinear registration had the greatest negative impact on natural variance in brain structure.

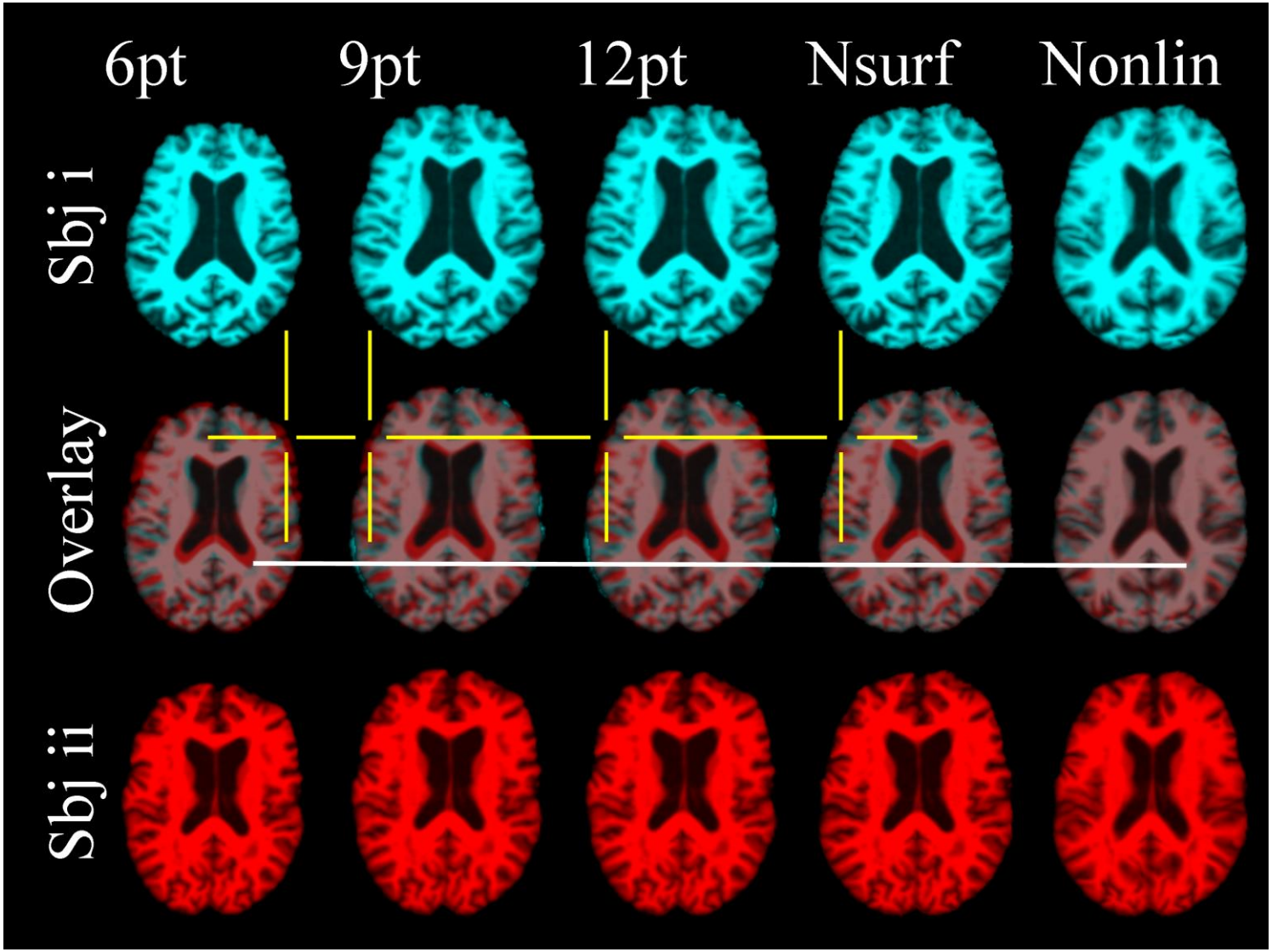


Figure 25. Comparison of registration methods
Methods with a low number of transformations (6 point, 9 point, and 12 point) did not always account for head size differences (indicated by the yellow crosshairs and bright cyan regions in the middle horizontal panel). Nonlinear (Nonlin) removed almost all of the within brain variance, e.g. lateral ventricle volume indicated by the white horizontal line in the middle panel. Nonlinear surface (Nsurf) registration adequately accounted for head size and maintained within brain variance (shown in the fourth vertical panel).

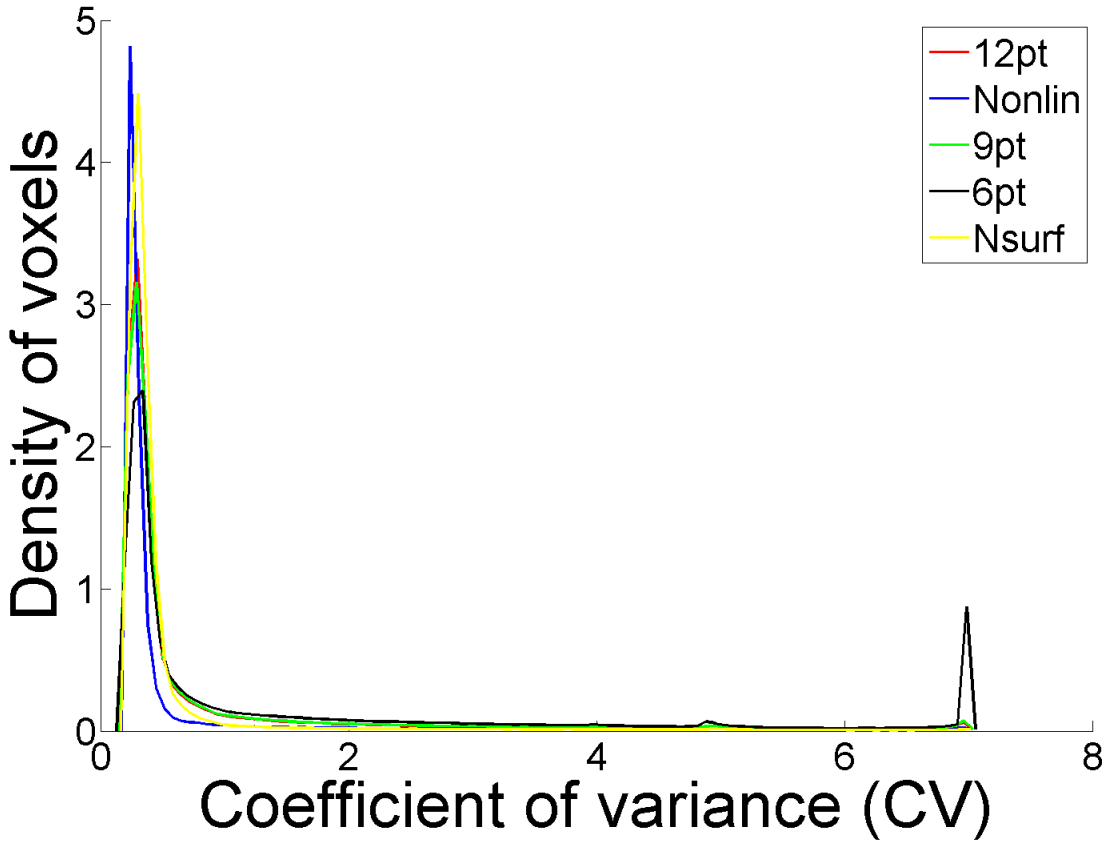


Figure 26. Histograms of coefficient of variance (CV) images from each registration method
The 9pt (green line) and 12pt (red line) registration methods were almost entirely overlapping in CV, with 12pt registration having a slightly lower median CV than 9pt (0.36 in 12pt versus 0.37 in 9pt).

Figure 26 shows the histograms of voxel variance (CV images) after linear, nonlinear, and nonlinear surface registration. It can be seen here that nonlinear registration has reduced the variance between subjects to almost zero in a large proportion of voxels.

I qualitatively observed that nonlinear surface registration more adequately accounted for head size differences. Although it appeared within brain structure was also maintained, I tested this quantitatively by computing proportional lateral ventricle volumes before and after registration. The error in proportional lateral ventricle volume following registration (median=2.76%, IQR=1.89%) indicated that nonlinear surface registration did adequately maintain within brain variance (Figure 27).

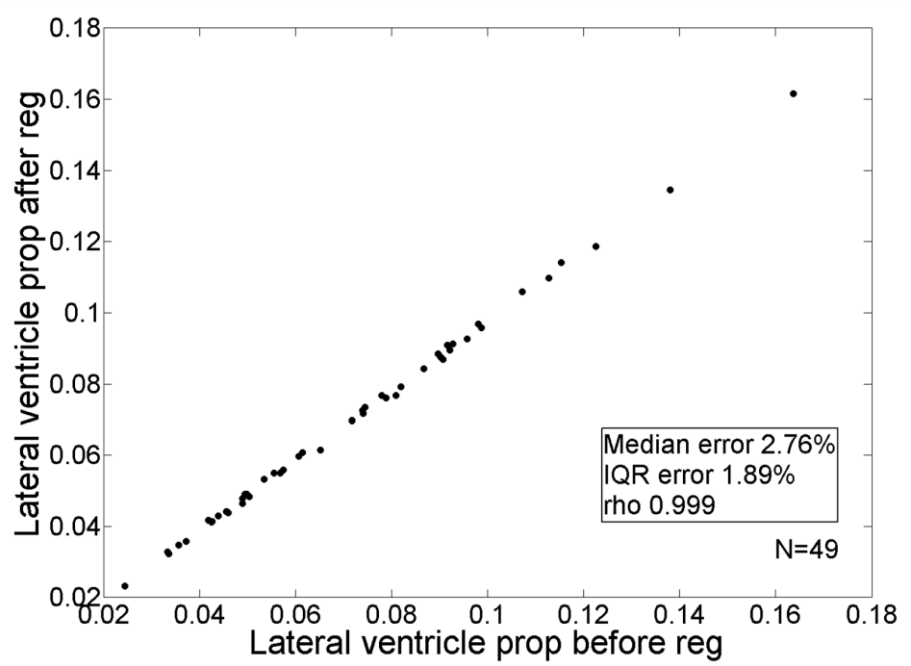


Figure 27. The effect of nonlinear surface registration on proportional lateral ventricle volume

Although the inner parts of the brain (e.g. lateral ventricles) were not unduly altered with nonlinear surface registration, the algorithm focused on the surface of the brain therefore may have been more likely to unduly affect cortical thickness. I tested this by computing cortical thickness before and after nonlinear surface registration in a group of AD subjects and a matched group of controls (the groups consisted of 89 subjects randomly selected from the ADNI cohort described in 5.2.3). Figure 28 shows that, while the algorithm increased cortical thickness (as expected given the increase in overall brain size; Figure 29), it did so to the same extent in both groups (ρ_{AD}=0.86, $\rho_{Control}$=0.85), i.e. it did not create a bias that would influence subsequent analyses.

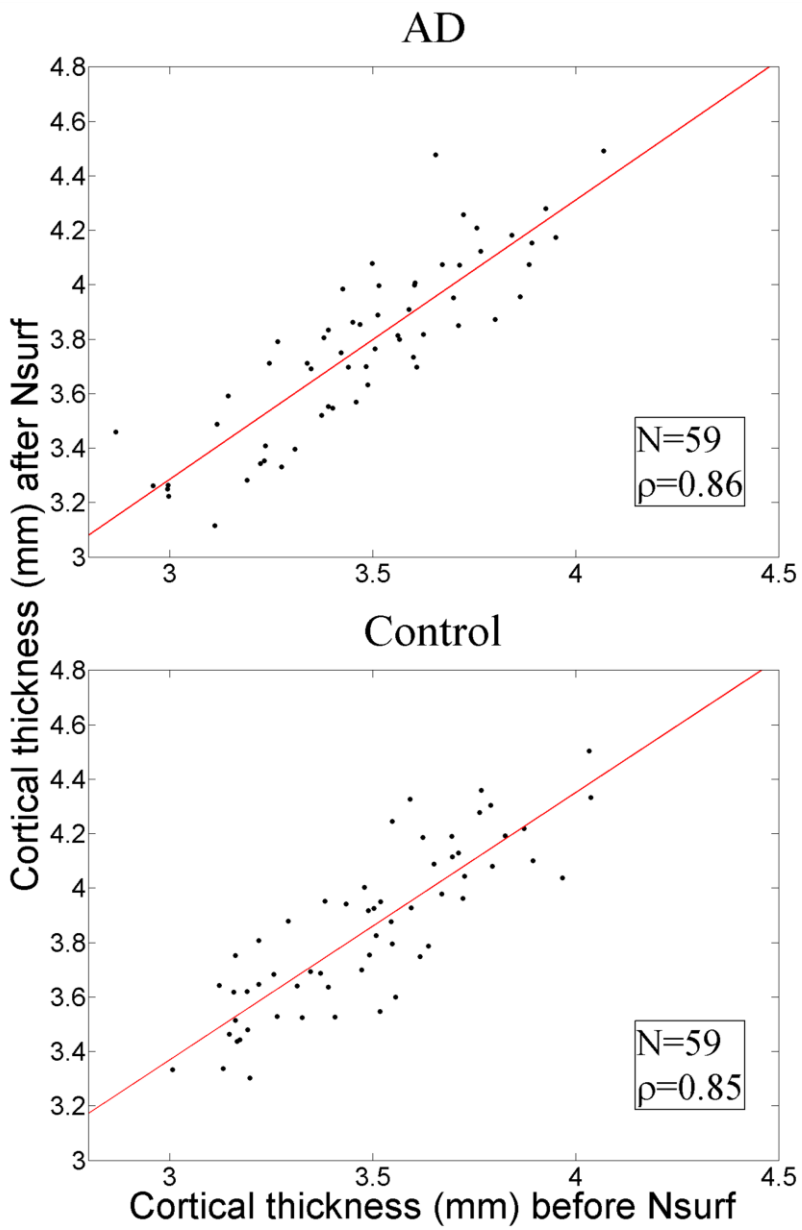


Figure 28. Cortical thickness before and after nonlinear surface (Nsurf) registration in a group of AD subjects (top panel) and a group of matched control subjects (bottom panel)
Cortical thickness was increased to the same level in both groups (as expected given the increase in overall brain size) and therefore Nsurf did not create a bias that would influence subsequent analyses.

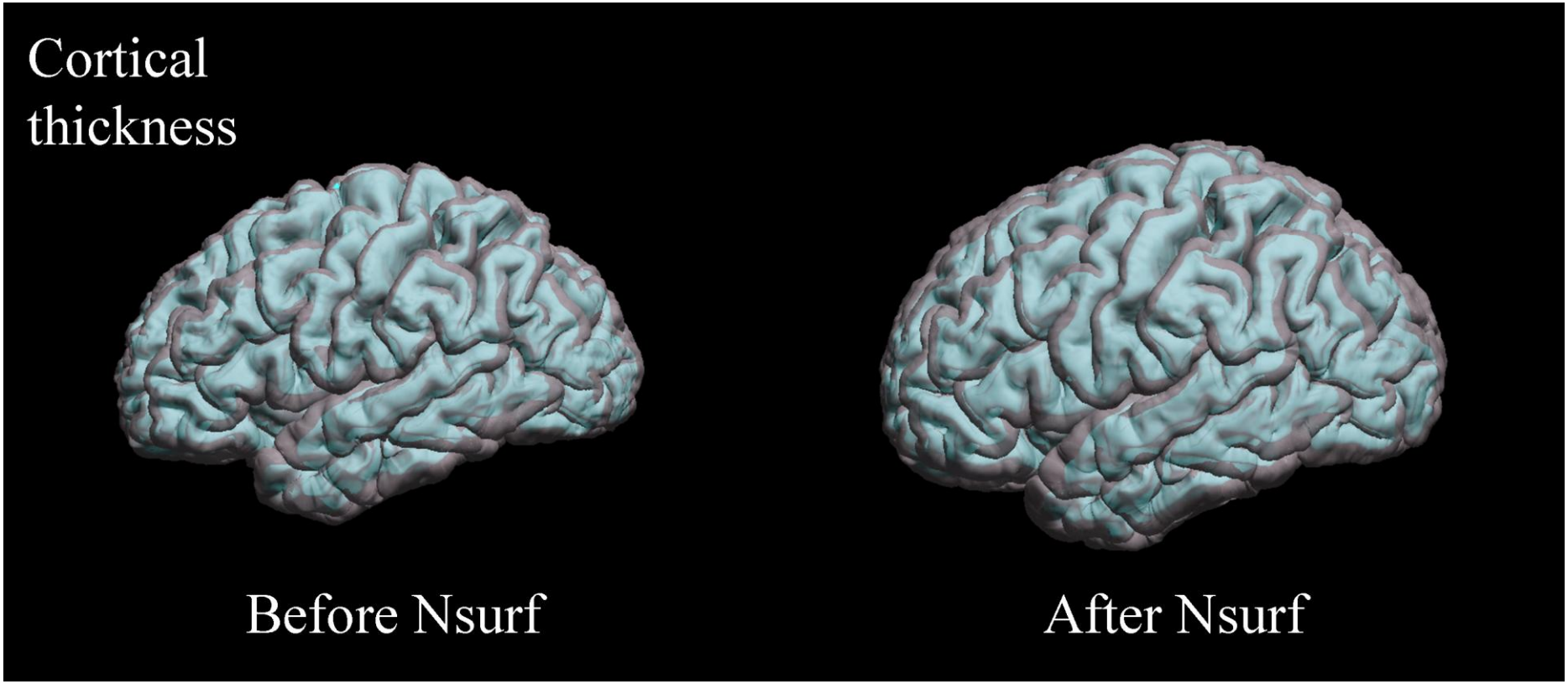


Figure 29. Cortical thickness before and after nonlinear surface registration
The increase in cortical thickness (shown in figure 28) after Nsurf is consistent with the increase in overall brain size shown here. The inner cortical boundary is shown in light blue and the outer cortical boundary is shown in dark pink.

5.5 Discussion

Because of the brain structure changes that occur through life (chapters 1-4), I had to assess specific image processing methods at different stages of the lifecourse. I discuss the implications of these assessments under each subject group heading, starting with infants.

5.5.1 Brain structure MRI processing in infants

Brain extraction in infants represented a unique challenge. Infant brains are much smaller than adult brains and therefore methods designed for use in adults were not entirely suitable. In particular, BET produced gross errors (that required substantial manual editing) despite extensive tweaking of the algorithms' parameters. This is not overly surprising and not a criticism of BET; it was simply not designed for use in infants. But given BETs' widely successful application, it was sensible to perform this test in infants and at least rule it out for future use. The automatic sequence of the BrainSuite pipeline produced smaller errors in some subjects, but still led to gross errors in others. Again, BrainSuite was not specifically designed for the nuances of the infant brain but did provide an incorporated manual editing platform to speed up processing time (by 10-15 minutes). While BrainSuite and BET were designed for use in adults, ANTS atlas-based segmentation was specifically designed for brains with challenging features, such as infants and older adults. It was not surprising then that ANTS, also with the longest computational processing time, produced the smallest errors and required the least manual editing time. However, it did still require manual editing and illustrates that in brains with such distinctive features, e.g. markedly smaller size, fully automatic processing may not be appropriate. The general shrinking of the brain in ageing meant that these issues were also present in aged adult subjects.

5.5.2 Brain structure MRI processing in aged adults

Brain extraction in aged adults was equally as challenging as in infant subjects. While the infant brain is growing, the aged adult brain is generally shrinking and this means that specifically designed programs, such as ANTS, are required to successfully brain extract these subjects. However, ANTS took up a large amount of computational time (~1 hour per subject) therefore I first attempted atlas-based segmentation using linear

registration (which has much less transforms and so takes much less time than ANTS). While it took dramatically less computational time (~2 minutes vs. ~1 hour), the limited number of transforms in linear registration drastically reduced the correspondence between the reshaped atlas and each subject. This meant that, while it saved on computational time, manual editing time was doubled with linear atlas-based segmentation. I therefore proceeded to use ANTS atlas based segmentation in aged adult subjects. This led to a significant reduction in errors but there were still mild to moderate errors in many subjects and these required approximately 10 minutes manual editing per subject. Moreover, there were gross errors in a very small number of subjects and these illustrate that, as in infants, fully automatic brain extraction may not be appropriate in aged adults.

Following brain extraction I automatically extracted GM, WM, and CSF images that may be used to sensibly compare subjects from different scanners and sequences. Especially for CSF, this process could be almost entirely automated even in these aged subjects (save for manual double-checking). However, some of these aged subjects had marked white matter hypointensities (potentially lesions) on their T1 images and these were sometimes incorrectly classified as GM. I required MR sequences that were not available for all subjects, e.g. FLAIR, T2*, to reliably correct these errors (Wardlaw et al., 2011). I therefore sought another method to allow sensible comparison of subjects across scanners and sequences.

Since I could not adequately address the issues with tissue segmentation (due to time and image sequence constraints), I implemented MR signal intensity standardisation (SIS). After addressing the issues of contrast alteration in the first version of the SIS algorithm, this appeared to bring the intensities of over 400 subjects with and without AD into a common framework. By computing GM before and after SIS it was apparent that the algorithm maintained the original anatomical structure of subjects. However, this validation is not definitive and I recommend that the results of SIS are thoroughly checked, both quantitatively and qualitatively, in the specific group of subjects under study.

While I recommend such meticulousness in all aspects of image processing in aged and infant subjects, the homogeneity of young to middle adult brains suggested a less rigorous approach could be adopted in these subjects.

5.5.3 Brain structure MRI processing in young to middle adults

The homogeneity between young to middle adult brains allowed me to test a downsampled brain extraction method. That is, I downsampled subjects to an atlas with lower resolution (number of slices) in the hope that this would negate or reduce manual editing time. While it did not completely negate the need, this method did reduce manual editing time by half (from 10 to 5 minutes). There were no gross errors in this method and even before implementation, manual editing in this group was the lowest across the lifespan. This suggests that the homogeneous anatomy of these subjects had a great deal to do with the success of this brain extraction method.

Following brain extraction, I extracted several regional brain volumes (hippocampus, parahippocampal gyrus, amygdala, putamen, caudate, and thalamus) also using atlas-based segmentation. Through visual assessment I found no gross errors in these segmentations. The segmentations were however not perfect and (due to time and resource constraints), I could not correct these small errors. This is a limitation common to many studies and an image processing method that is commonly applied (Maldjian et al., 2003). Therefore, since I am interested in testing commonly applied brain imaging statistical methods, it is arguably only a minor limitation for my purposes. Future studies with manually edited regional volumes will assess the effect of this limitation in determining true measures across the lifecourse.

Although extraction of brain image volumes requires specific methods given the age of subjects, registration (spatial normalisation) is generally applicable across the lifecourse.

5.5.4 Registration

Linear registration only alters brains via their main axes therefore does not unduly alter within brain structure such as lateral ventricle volumes, i.e. lateral ventricle volume is scaled according to the whole brain maintaining its original ratio. This means that spurious differences in head size that occur in regions other than the main axes may remain. Conventional nonlinear registration addresses this problem however it does not maintain within brain structure variance, e.g. lateral ventricle volumes are warped to those of the atlas/ registration target. I therefore devised Nsurf registration to fully account for head size differences but maintain within brain structure variance of interest. Validation of this method in aged subjects suggested that it sufficiently

addressed these issues without creating systematic bias between groups, e.g. in cortical thickness between AD subjects and controls. However, it should be noted that, as with any registration method, Nsurf does alter overall brain structure, e.g. Figure 29. Moreover, my validation only applies to the limited number (n=118) of ageing and AD subjects I assessed. Because of this I recommend that Nsurf is appropriately validated before it is used in subjects at other stages of the lifecourse.

5.5.5 Summary

The changing nature of the brain through life requires image processing methods that are specific to the age of subjects under study. Different methods may also be adopted depending on the purpose of a study, e.g. in a functional study investigators may wish to entirely remove brain structure variance between subjects by registering each subject to an atlas with a very high number of transforms. This work was interested in quantifying normal brain variance that may be implicit in subsequent neurological disease. I therefore devised a registration method (Nsurf) that removed spurious differences, e.g. head size differences, while it maintained variance that may indicate disease, e.g. lateral ventricle volume.

Having extracted relevant data in a manner most suitable for my purposes, I proposed a novel statistical analysis method for defining the normal limits of brain tissue volumes; this method is described in the next chapter.

6. A novel statistical method for definition of the distributions and boundaries of ageing brain volumes[4]

6.1 Introduction

Methods to define normal brain volumes across age are often based on central tendency, e.g. mean or median, statistics. The review in chapter 4 found that, for over 20 years, mean-based statistical models have been used extensively to define normal brain volumes across age (see chapter 4 for the full review or Table 17 for a summary).

Mean estimates may be extrapolated to define the distributions and clinical boundaries of brain volumes if the variance in brain volumes is equally Gaussian between ages and disease states. However, if the variance is not equally Gaussian, the distribution of differences in brain volumes between ages may not be well represented by models based on mean estimates (Elveback et al., 1970; Koenker and Bassett Jr, 1978; Freedman et al., 2007). That is, the true range of normality and boundaries with pathology may be obscured if extrapolated from these models.

Several independent studies have shown that variance in brain tissue volume is unequal between ages (Manolio et al., 1994; Ge et al., 2002; Resnick et al., 2003; Kruggel, 2006; Farrell et al., 2009), i.e. the range of volumes generally increases with age. However, data transformation methods were previously used in brain image analyses, e.g. Box-Cox (Box and Cox, 1964; Kruggel, 2006), and these removed much of this inequality to fit the requirements of models based on mean estimates. While this approach is useful in research to identify general patterns of brain ageing, it may mask the true boundaries of normality and the subtle early signs of disease. I found no other previous study that attempted to define the distribution and boundaries of normal brain volume differences between ages. The true distribution of differences in brain tissue volume between ages is therefore largely unknown.

---

[4]This was supported by a travel grant award from Guarantors of Brain (www.guarantorsofbrain.org.uk) and was published/ presented as follows:

Dickie, D.A., Job, D.E., Rodríguez González, D., Shenkin, S.D., Ahearn, T.S., Murray, A.D., Wardlaw, J.M. (2013). Variance in brain volume with advancing age: implications for defining the limits of normality. PLoS ONE 8(12): e84093.
Dickie, D.A., Job, D.E., Rodríguez González, D., Shenkin, S.D., Wardlaw, J.M. (2012). A databank, rather than statistical, model of normal ageing brain structure to indicate pathology. *Proceedings of the 18th Annual Meeting of the Organization for Human Brain Mapping*. Beijing, China

Table 17. Methods to define normal brain volumes across age

Study	No. of Subjects	Age range[a]	Mean ±SD Age[a]	Statistical Method[b]
Allen et al., 2005	87	22.0–88.0	49.4 ± 20.8	**Multiple regression**
Courchesne et al., 2000	116	1.6–80.0	21.4 ±20.0	**Regression analyses**
DeCarli et al., 2005	2081	34.0–96.0	62.4 ±10.4	**Linear regression**
Fotenos et al., 2005	94	65.0–95.0	78.0 ±8.0	**Hierarchical polynomial regression**
Ge et al., 2002	54	20.0–86.0	46.8 ±19.3	**Least–squares regression**
Giorgio et al., 2010	66	23.0–81.6	36.7[c]	**Regression analysis**
Good et al., 2001	465	17.0–79.0	29.5[d]	**General Linear Model**
Gur et al., 1991	69	18.0–80.0	41.4 ±20.2	**Multivariate analysis of variance**
Jernigan et al., 2001	78	30.0–99.0	64.0 ±17.4	Nonparametric monotone regression
Kruggel, 2006	502	16.0–70.0	30.0 ±9.6	**Linear, quadratic regression**
Raz et al., 2005	72	20.0–77.0	52.6 ±14.1	**Latent difference model**
Sowell et al., 2003	176	7.0–87.0	32.4 ±21.8	**Quadratic multiple regression**
Walhovd et al., 2005	25	67.0–88.0	74.3 ±4.8	**Regression analyses**
Ziegler et al., 2011	547	19.0–86.0	48.1 ±16.6	**General linear model**

Note: Bolded entries show methods based on means; [a]In years; [b]This is the statistical method used to provide the majority of results, recorded as stated in the corresponding manuscript; [c]Median; [d]Estimated median; SD=standard deviation.

Specifically, as they have not previously been calculated (Table 17), it is not known whether percentile rank differences are equal to mean (parametric) differences in brain volume. In parametric regression models it is implicitly assumed that mean differences approximate percentile rank differences (Freedman et al., 2007). Percentile ranks are levels that represent percentages of subjects within a distribution, e.g. the bottom 5% of subjects in a distribution have a value equal to or less than the 5th percentile rank value (Freedman et al., 2007). To define these levels with mean-

based models one must assume, for example, that the differences in brain tissue volume between 60 and 70 year olds at lower ranks, e.g. the 5th percentile, are the same as the differences between 60 and 70 year olds at higher ranks, e.g. the 95th percentile. It is unknown whether or not this is true. If it is not true then potentially important biological information may be lost by extrapolation from mean-based models. In particular, the limits of normality and boundaries with pathology may be misrepresented (Elveback et al., 1970; Freedman et al., 2007).

In this chapter, using publicly available brain images, I attempted to determine whether mean estimates of brain volumes across age approximated percentile rank estimates.

6.2 Methods

6.2.1 Subjects

MR brain images from 137 "normal" subjects (n=60, 44% female), mean age ~76 (70-89) years, were acquired from the ADNI databank (Table 15, page 71). These subjects did not have dementia, but potentially had non-debilitating conditions common in ageing, e.g. hypertension. A further 90 subjects (n=65, 72% female) with similar clinical characteristics and a mean age of ~75 (60-89) years, were acquired from the OASIS databank (Marcus et al., 2007c) (Table 15, page 71). According to the systematic review in chapter 2, ADNI and OASIS were the only public sources of structural MR brain images that had medical and cognitive metadata representative of the characteristics of normal older people (≥60 years). These subjects were reported to be representative of normal older people through cognitive and physical tests but actual measures of potentially confounding variables, e.g. blood pressure, for all subjects were not available to me (Marcus et al., 2007c). I combined the ADNI and OASIS samples to create a single normal subject sample of n=227 subjects.

A sample of MR brain images from 124 subjects diagnosed with AD (n=58, 47% female) and a mean age of ~75 (55-89) years was also acquired from ADNI (Table 15, page 71). A further 95 AD subjects with similar demographics were acquired from OASIS (Table 15, page 71) and combined with the ADNI subjects to create a single AD sample of n=219 subjects.

6.2.2 MR brain image processing

Non–brain structure was removed from the images and grey matter, white matter, and cerebrospinal fluid volumes were calculated by the steps described in chapter 5. The advanced age of these subjects meant that many of them had what appeared to be WM lesions (hypointense WM regions on T1-wieghted images that are hyperintense on T2-weighted images). FAST sometimes incorrectly classified these regions as GM. I did not have the MRI sequences, e.g. fluid attenuated inversion recovery (FLAIR), required to accurately address these errors for all subjects (Marcus et al., 2007c). Although incorrectly classified as GM, these regions were still correctly classified as tissue, i.e. not CSF. I therefore added the GM and WM volumes within each subject to calculate overall brain tissue volume in voxels. Brain tissue volumes were then normalised (divided) by total intracranial volume (TIV=tissue+CSF). I do not report normalised CSF regression models because they were the exact inverse of normalised whole brain tissue volume models.

6.2.3 Calculating mean and percentile rank regression models of brain volume by age

When age was expressed in one year intervals, there were very few subjects at some ages, meaning that it was not possible to calculate the values of percentile ranks at these ages. Therefore, I expressed age in the following intervals: <70, 70-74, 75-79, 80-84, and 85-89 years.

Mean differences in brain tissue volume between age groups were calculated in each sample with "PROC REG" in the Statistical Analysis System (SAS) v9.3 (http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#reg_toc.htm). PROC REG produces a regression equation ($y=\beta x_1 \ldots \beta x_n+c$) to describe mean differences in a dependent variable (y), e.g. brain tissue volume, between values of an independent variable (x), e.g. age group. The beta (β) coefficient of this equation defines the size of mean differences between groups.

The 5th, 25th, 50th, 75th, and 95th percentile ranks of brain volume were directly calculated for each age group by equation 4,

$$np = j + g \qquad (4.1)$$

$$y = 1/2(x_j + x_{j+1}) \quad \text{if } g = 0 \qquad (4.2)$$

$$y = x_{j+1} \quad \text{if } g > 0$$

where n is the number of subjects, for the tth percentile $p=t/100$, j is the integer part of np, g is the fractional part of np, y is the tth percentile, and x_1, x_2, ... , x_n are the ordered values of brain tissue volume. Differences in volume between ages (regression equations) at these percentile ranks (rather than the mean) were then calculated with PROC QUANTREG in SAS.

For both mean and percentile rank estimates, I initially performed linear regressions. These were used to produce residual plots (actual minus regression predicted volumes by age group). The linearity assumption is in question if these plots show a systematic pattern, e.g. particular age groups with skewed positive or negative residuals (Freedman et al., 2007). When the linearity assumption was in question, I performed nonlinear (cubic) regression to define differences between ages (Fotenos et al., 2005). Cubic regression was chosen because previous literature indicated brain tissue exhibits a nonlinear decline in ageing (>60 years) but not a nonlinear decline with subsequent regrowth (i.e., it does not exhibit a higher order polynomial pattern; Good et al., 2001; Fotenos et al., 2005).

The representativeness of mean estimates was illustrated by relative percentage error between the expected (mean) and observed (percentile rank) regression predictions, calculated by equation 5,

$$\frac{\hat{y}_{\mu p,i} - \hat{y}_{dp,i}}{\Delta_\mu} \times 100 \qquad (5)$$

where $\hat{y}_{\mu p,i}$ is the mean-based prediction of percentile rank p for age group i, $\hat{y}_{dp,i}$ is the directly calculated percentile rank prediction, and Δ_μ is the overall mean change in brain tissue volume (mean regression predicted volume at <70 years minus mean regression predicted volume at 85-90 years). Due to the potential for subtle

differences between ages and disease (Fotenos et al., 2005; Farrell et al., 2009), results are reported with four significant figures.

6.3 Results

6.3.1 Mean and variance of brain volume within the normal and AD samples

The mean and SD of normalised brain volume in each age group in the normal and AD samples are shown in Table 18. Given the potential for subtle differences between age groups, volumes are given to four significant figures. Variance in brain tissue volume generally increased with age in the normal sample but decreased with age in the AD sample.

Table 18. The mean and standard deviation of normalised brain volume in each age group in the normal and AD samples

Sample	Age	n	Mean	SD
Normal	<70	25	0.7555	0.0130
	70-74	89	0.7569	0.0196
	75-79	56	0.7490	0.0204
	80-84	34	0.7418	0.0193
	85-89	23	0.7359	0.0210
AD	<70	43	0.7399	0.0236
	70-74	56	0.7350	0.0208
	75-79	49	0.7330	0.0209
	80-84	49	0.7292	0.0199
	85-89	22	0.7128	0.0187

Note: SD=standard deviation.

Residual plots (actual minus regression predicted volumes by age group) from the mean linear regressions in each sample are shown in Figure 30. The different spread of points at each age group further illustrates unequal variance between ages. Moreover, skewed residuals at 70 years in the normal sample (few positive compared to negative residuals) and a similar pattern at 85-90 years in the AD sample suggested that the linearity assumption was in question. I therefore performed nonlinear regression of brain tissue volume across age in each sample.

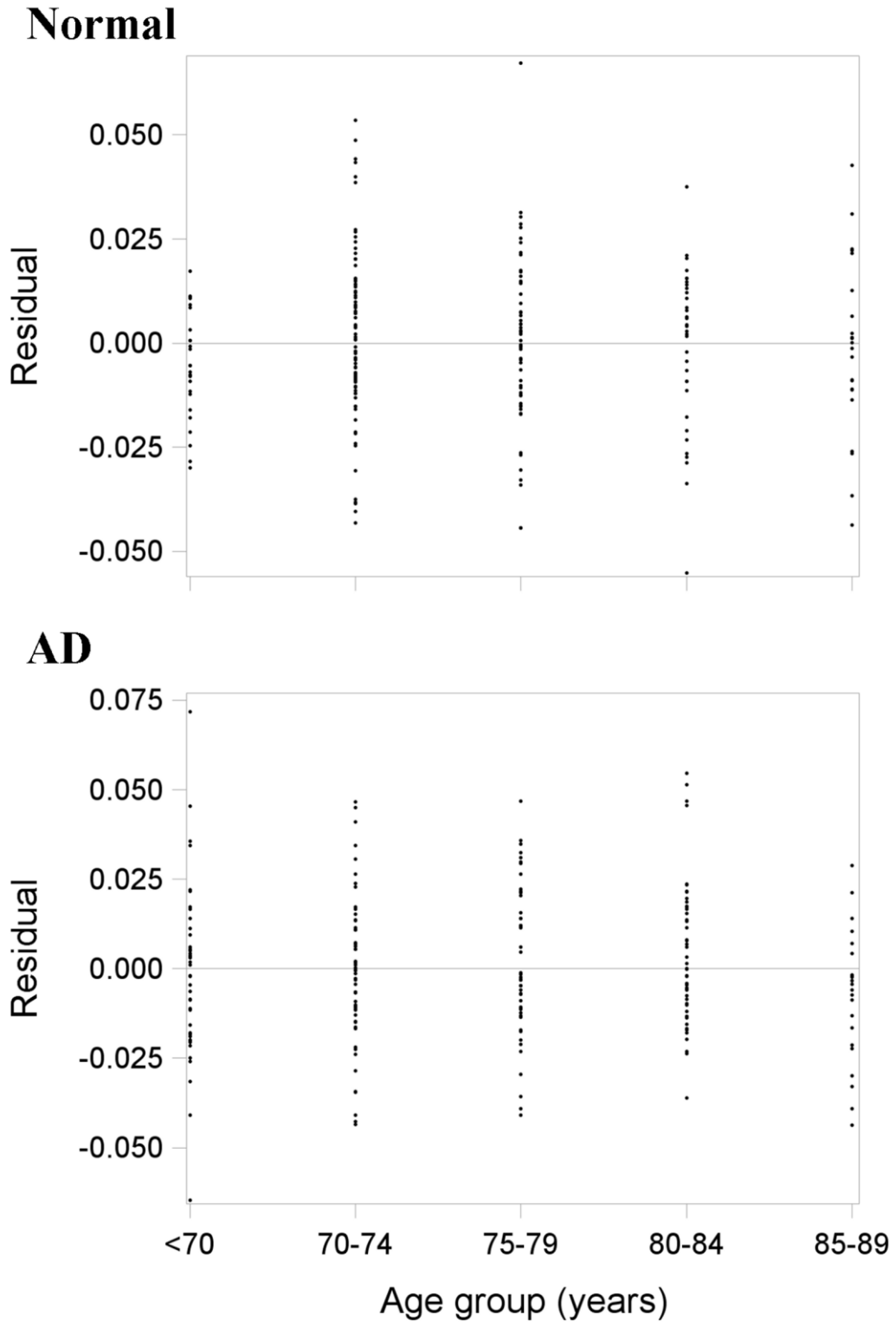


Figure 30. Residual plots (actual minus mean linear regression predicted brain volumes by age) in the normal (top panel; n=227) and AD (bottom panel; n=219) samples.
There are skewed residuals at 70 years in the normal sample (top) and a similar pattern at 85-90 years in the AD sample (bottom). This means that the linearity assumption was in question.

6.3.2 Mean and percentile rank regression estimates of brain volume across age within the normal sample

Mean and percentile rank regression estimates of brain volume across age in the normal sample are listed in Table 19 and illustrated in Figure 31 and Figure 32. As the dashed (percentile rank) lines within each graph (Figure 31) are generally not parallel, the distribution of differences in brain tissue volume between ages was not well represented by mean estimates. The diverging percentile rank lines further illustrate that variance in brain tissue volume increased with age in normal subjects.

Mean estimates generally overestimated differences in brain tissue volume between ages at percentile ranks above the mean, i.e. in the upper percentile ranks, mean regression beta underestimated brain tissue volume at advanced ages. For example, the mean-based prediction of the 95th percentile of brain volume at 85-90 years was short by 74% of the overall expected change between <70 and 90 years (Table 19). This is in contrast to percentile ranks below the mean, where mean estimates generally underestimated differences between ages, i.e. in the lower percentile ranks, mean regression beta overestimated brain tissue volume at advanced ages. For example, the mean-based prediction of the 5th percentile of brain volume at 85-90 years was inflated by 66% of the overall expected change between <70 and 90 years (Table 19).

Table 19. Mean and percentile rank regression estimates of normalised brain tissue volume across age in the normal sample

Rank	Age (x)	c	Beta_x	Beta_x^2	p prediction	μ prediction (Δ_μ)	% error
MEAN		0.7584	0.0007	-0.0011			
5th		0.7397	-0.0089	0.0002			
	<70 (1)				0.7310	0.7393	35
	70-74 (2)				0.7227	0.7367	59
	75-79 (3)				0.7148	0.7319	72
	80-84 (4)				0.7073	0.7249	75
	85-89 (5)				0.7002	0.7157 (0.0236)	66
25th		0.7545	-0.0049	-0.0002			
	<70 (1)				0.7494	0.7541	20
	70-74 (2)				0.7439	0.7515	32
	75-79 (3)				0.7380	0.7467	37
	80-84 (4)				0.7317	0.7397	34
	85-89 (5)				0.7250	0.7305 (0.0236)	23
50th		0.7533	0.0036	-0.0014			
	<70 (1)				0.7555	0.7529	-11
	70-74 (2)				0.7549	0.7503	-19
	75-79 (3)				0.7515	0.7455	-25
	80-84 (4)				0.7453	0.7385	-29
	85-89 (5)				0.7363	0.7293 (0.0236)	-30
75th		0.7674	0.0038	-0.0016			
	<70 (1)				0.7696	0.7670	-11
	70-74 (2)				0.7686	0.7644	-18
	75-79 (3)				0.7644	0.7596	-20
	80-84 (4)				0.7570	0.7526	-19
	85-89 (5)				0.7464	0.7434 (0.0236)	-13
95th		0.7747	0.0067	-0.0016			
	<70 (1)				0.7798	0.7743	-23
	70-74 (2)				0.7817	0.7717	-42
	75-79 (3)				0.7804	0.7669	-57
	80-84 (4)				0.7759	0.7599	-68
	85-89 (5)				0.7682	0.7507 (0.0236)	-74

Note: This table shows the regression equations for the mean and each percentile rank. Percent errors between predictions are calculated for each age relative to the overall mean regression change (Δ_μ=mean regression predicted volume at <70 years minus mean regression predicted volume at 85-90 years). Positive percent errors indicate that the mean regression underestimated differences between ages, i.e. overestimated brain tissue volume at advanced ages. Negative percent errors indicate that the mean regression overestimated differences between ages, i.e. underestimated brain tissue volume at advanced ages. Age groups were coded as 1 (<70) to 5 (85-89); c=intercept; μ=mean; p=percentile rank.

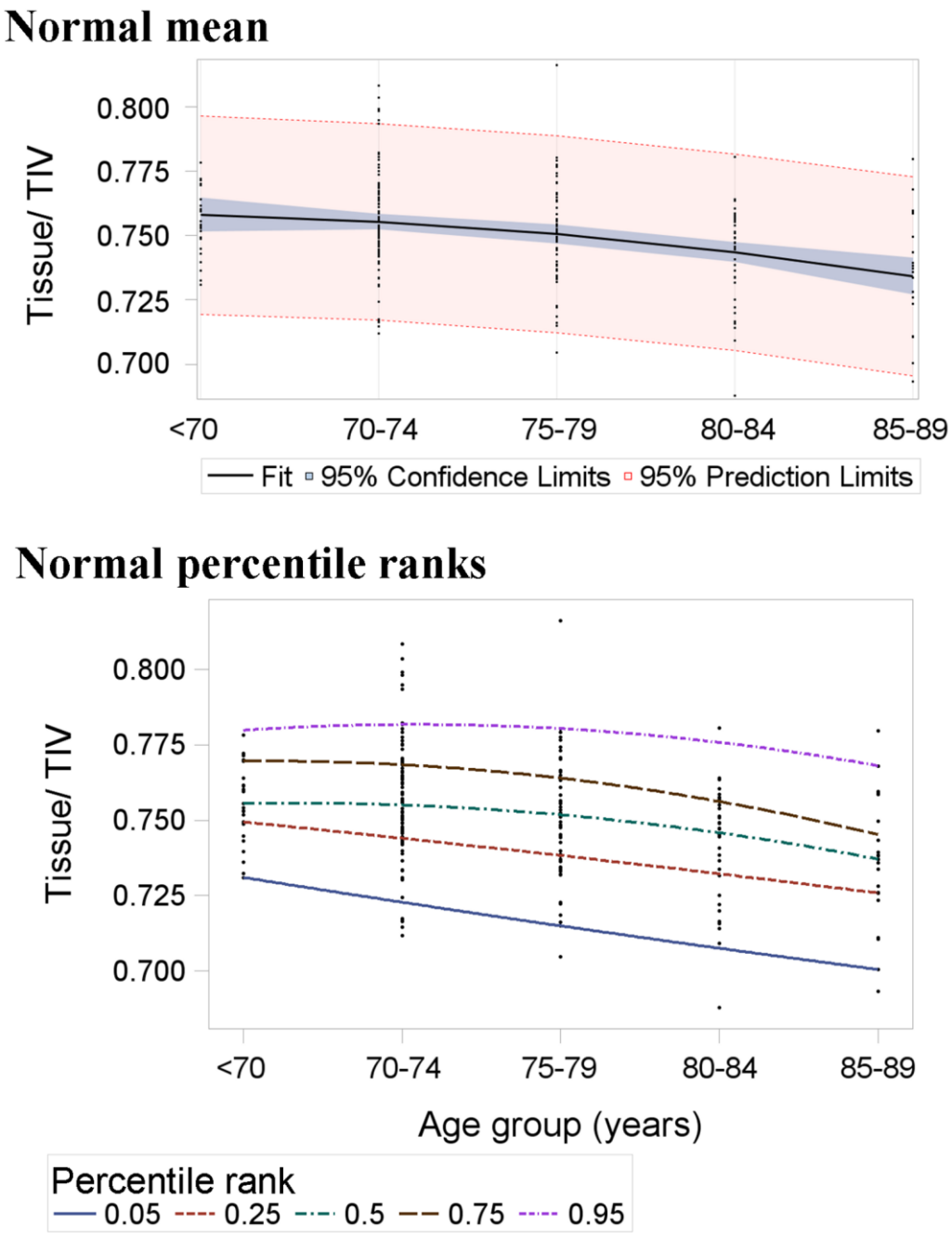


Figure 31. Mean (top panel) and percentile rank (bottom panel) regression estimates of brain tissue volume across age in the normal sample (n=227).
The slopes of these lines represent the beta coefficients in table 4. The mean-based model expects all percentile ranks to change at the same rate, i.e. be parallel. The diverging percentile ranks show that this is not the case and that variance in brain volume generally increased with age in the normal subjects. Although some may appear linear, each line is the result of nonlinear regression (table 4).

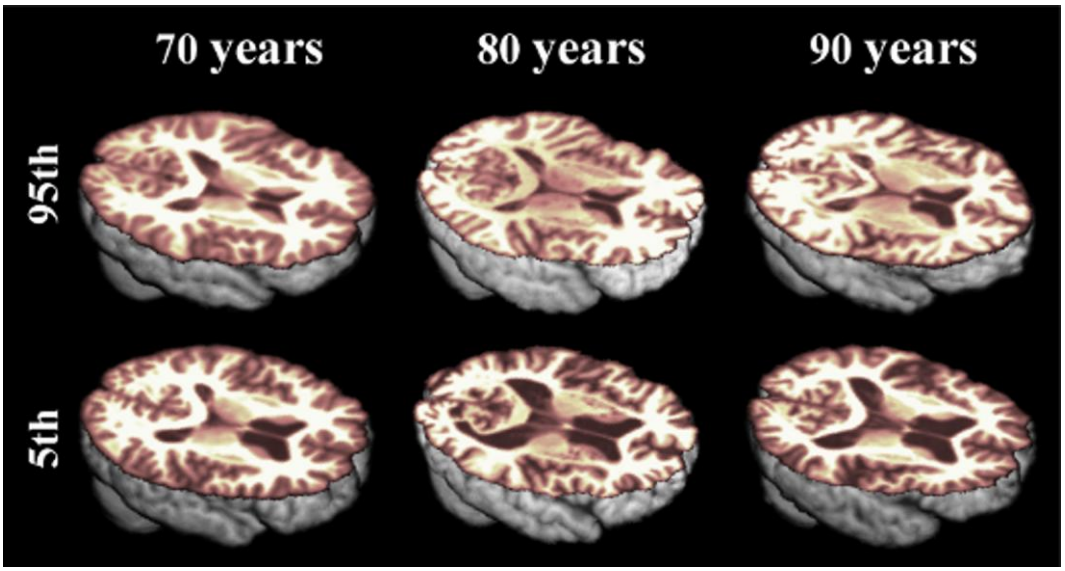


Figure 32. Illustration of the varying differences in normal ageing brain tissue volume, according to percentile rank.
There were much greater differences between ages at the 5th percentile of brain tissue volume (bottom panel) than between ages at the 95th percentile (top panel) of normal subjects.

6.3.3 Mean and percentile rank regression estimates of brain volume across age within the AD sample

Mean and percentile rank regression estimates of brain volume across age in the AD sample are listed in Table 20 and illustrated in Figure 33. As in normal subjects, the lack of parallel lines (Figure 33) shows that the distribution of differences in brain tissue volume between ages in AD was not well represented by mean estimates. The converging percentile rank lines further illustrate that variance in brain tissue volume decreased with age in AD subjects.

Opposite to the normal sample, mean estimates generally underestimated differences in brain tissue volume between ages at percentile ranks above the mean, i.e. in the upper percentile ranks, mean regression beta overestimated brain tissue volume at advanced ages. For example, the mean-based prediction of the 95th percentile of brain volume at 85-90 years was inflated by 47% of the overall expected change between <70 and 90 years (Table 20). This is in contrast to percentile ranks below the mean, where mean estimates generally overestimated differences between ages, i.e. in the lower percentile ranks, mean regression beta underestimated brain tissue volume at advanced ages. For example, the mean-based prediction of the 5th

percentile of brain volume at 85-90 years was short by 28% of the overall expected change between <70 and 90 years (Table 20).

Table 20. Mean and percentile rank regression estimates of normalised brain tissue volume across age in the AD sample

Rank	Age (x)	c	Beta_x	Beta_x^2	p prediction	μ prediction (Δ_μ)	% error
MEAN		0.73638	0.0035	-0.0015			
5th		0.6950	0.0072	-0.0020			
	<70 (1)				0.7002	0.6970	-14
	70-74 (2)				0.7014	0.6960	-24
	75-79 (3)				0.6986	0.6920	-30
	80-84 (4)				0.6918	0.6849	-31
	85-89 (5)				0.6810	0.6749 (0.0222)	-28
25th		0.7186	0.0062	-0.0019			
	<70 (1)				0.7229	0.7206	-10
	70-74 (2)				0.7234	0.7196	-17
	75-79 (3)				0.7201	0.7156	-20
	80-84 (4)				0.7130	0.7085	-20
	85-89 (5)				0.7021	0.6985 (0.0222)	-16
50th		0.7474	-0.0060	0.0000			
	<70 (1)				0.7414	0.7494	36
	70-74 (2)				0.7354	0.7484	59
	75-79 (3)				0.7294	0.7444	68
	80-84 (4)				0.7234	0.7373	63
	85-89 (5)				0.7174	0.7273 (0.0222)	44
75th		0.7399	0.0098	-0.0025			
	<70 (1)				0.7472	0.7419	-24
	70-74 (2)				0.7495	0.7409	-39
	75-79 (3)				0.7468	0.7369	-45
	80-84 (4)				0.7391	0.7298	-42
	85-89 (5)				0.7264	0.7198 (0.0222)	-30
95th		0.7794	-0.0016	-0.0009			
	<70 (1)				0.7769	0.7814	20
	70-74 (2)				0.7726	0.7804	35
	75-79 (3)				0.7665	0.7764	45
	80-84 (4)				0.7586	0.7693	48
	85-89 (5)				0.7489	0.7593 (0.0222)	47

Note: This table shows the regression equations for the mean and each percentile rank. Percent errors between predictions are calculated for each age relative to the overall mean regression change (Δ_μ=mean regression predicted volume at <70 years minus mean regression predicted volume at 85-90 years). Positive percent errors indicate that the mean regression underestimated differences between ages, i.e. overestimated brain tissue volume at advanced ages. Negative percent errors indicate that the mean regression overestimated differences between ages, i.e. underestimated brain tissue volume at advanced ages. Age groups were coded as 1 (<70) to 5 (85-89); c=intercept; μ=mean; p=percentile rank.

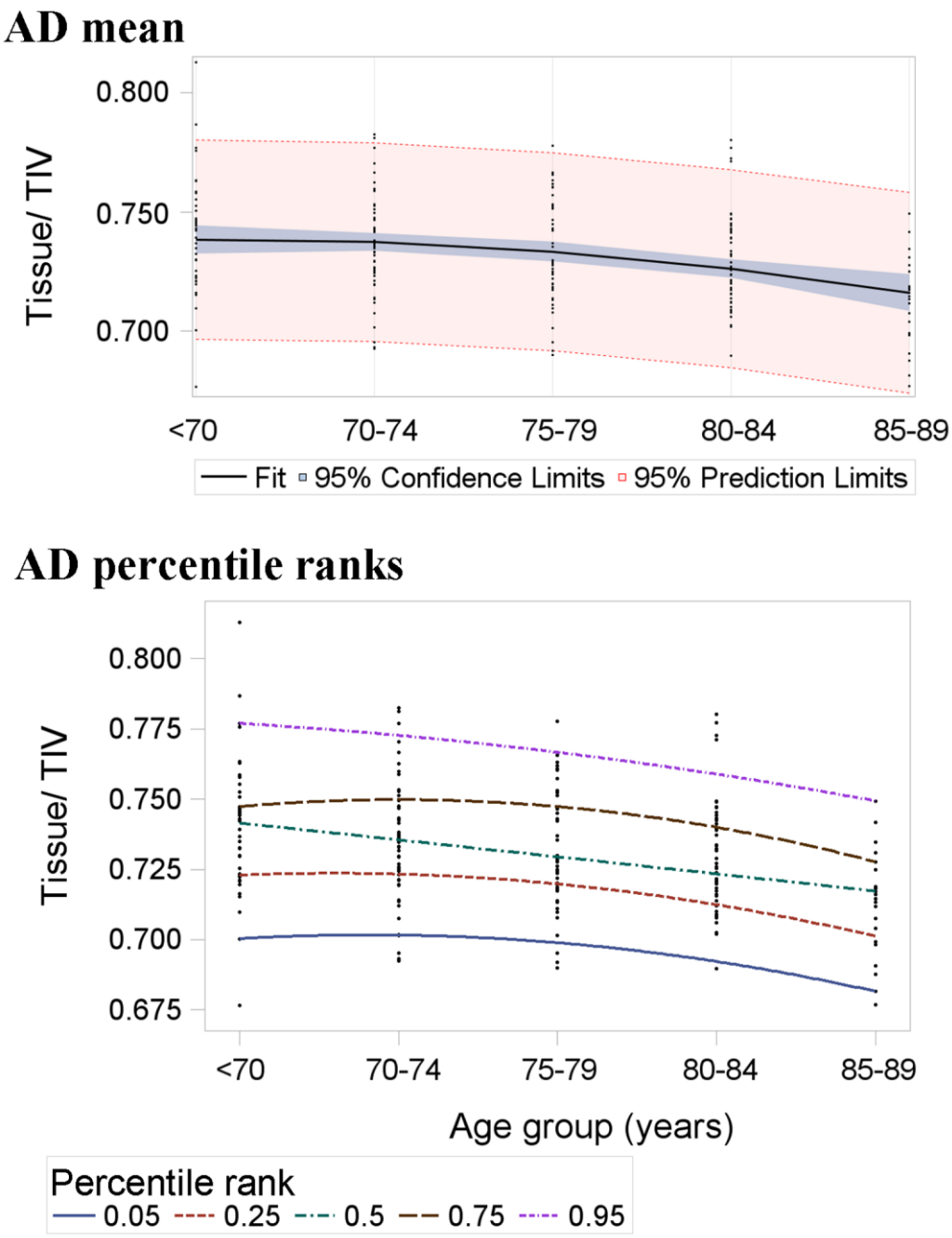


Figure 33. Mean (top panel) and percentile rank (bottom panel) regression estimates of brain tissue volume across age in the AD sample (n=219).
The slopes of these lines represent the beta coefficients in table 5. The mean-based model expects all percentile ranks to change at the same rate, i.e. be parallel. The converging percentile ranks show that this is not the case and variance in brain volume generally decreased with age in the AD subjects. Although some may appear linear, each line is the result of nonlinear regression (table 5).

6.3.4 Comparison of normal and AD samples

There was increased overlap between normal and AD brain tissue volumes with advancing age, as illustrated by the diverging normal subject percentile ranks and

converging AD subject percentile ranks (Figure 31 and Figure 33). Differences between normal ageing subjects at lower percentile ranks were similar to, or even greater than mean differences between AD subjects. For example, 5th percentile rank differences in normal subjects were 38.7% greater than mean differences in AD subjects.

6.4 Discussion

I have shown in the publicly-available data that variances in brain tissue volume are unequal across older ages. As a result of this, the distributions and clinical limits of brain tissue volume may not be well represented by mean estimates. Differences in brain tissue volume between normal subjects at the lowest percentile ranks (below the mean) were considerably greater than differences at the highest percentile ranks (above the mean). For example, mean estimates inflated volumes by 35 to 75% at the 5th percentile of normal subjects whereas they were short by 23 to 74% at the 95th percentile of normal subjects. This normal variation needs to be adequately defined so that it is not incorrectly attributed to neurodegenerative disease. While I had insufficient data to define true population variance, this proof of concept study suggests that percentile rank statistical models will be required to do so.

Statistical models of the normal ageing brain may be used to support earlier diagnoses of AD and related disorders (Fox et al., 2001; Farrell et al., 2009; McEvoy et al., 2009). Models based on mean estimates found that differences in brain tissue volume between ages were greater in AD than in normal ageing (Fotenos et al., 2005). However, these models assume that the distributions of normal and abnormal brain volumes are equally Gaussian within age, and that the overlap between these distributions does not change with age (Freedman et al., 2007). I found that the distributions of brain volume were not equal within age and that the overlap between normal and AD brain volumes increased with advancing age.

Further, I found that differences in brain tissue volume between normal ageing subjects at lower percentile ranks may be similar to, or even greater than differences between patients diagnosed with AD. Therefore, if a group of subjects acquired for controls in a clinical trial was unknowingly skewed to lower percentiles, true treatment effects in brain volume between normal ageing and AD may be obscured.

A percentile rank-based reference for brain volumes may then be useful to quantitatively rank individuals or to determine if the distribution of a control group is skewed. Given the wide and irregular variance in brain volume that I have identified in a relatively small number of apparently normal subjects, this percentile rank-based reference will require much more data than are publicly available at present (Dickie et al., 2012).

New databanks such as Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD), which provides longitudinal data from 23 cognitively tested normal ageing subjects (Malone et al., 2013), may help to address this shortage. The Australian Imaging Biomarkers & Lifestyle Flagship Study of Ageing (AIBL) databank is similar to ADNI and OASIS and includes 177 control subjects aged over 60 years. However, these subjects are not generally representative of the normal ageing population as they were preferentially selected as APOE ε4 allele carriers (Ellis et al., 2010).

Other neuroimaging projects are ongoing or initiating that may, in due course, provide the required data. For example, we are building a brain image databank to provide a normal reference for the brain using existing data from >1000 cognitively tested normal older subjects aged mostly between 55 and >90 years (http://www.sinapse.ac.uk/research-resources/brains-project). These data are in the process of being collated and were not available at the time of this study. In the future, by iteratively adding subjects and monitoring the subsequent fluctuations in percentile rank values, I may determine the amount of data required to create robust models that represent the true distributions of brain volumes in normal ageing and disease.

The limited number of subjects available to the present study (n=446) meant that I had to express age in five year rather than one year intervals. Differences between subjects within these intervals could not then be calculated here but could be calculated in future studies with larger samples. The age groups in this study did not have equal sample sizes and this may have contributed to the unequal variance in brain volumes (Freedman et al., 2007). However, a larger number of subjects generally leads to greater variance (Freedman et al., 2007). Therefore, since sample size generally decreased while variance increased with age in the normal subjects, future studies with more subjects may further illustrate our point that variance in the brain increases during normal ageing. This further suggests that percentile rank

models may be required to adequately describe the "true" levels and limits of normal brain structure.

While I have shown that percentile rank analyses will likely be required to define the range of ageing brain structure, it may be that the wide variance described here (e.g., Figure 33) is due to the limited number of subjects currently available. Future work with much larger samples will be required to determine whether percentile ranks vary as greatly in the wider normal ageing population.

The inclusion of "normal" subjects with and without hypertension or other risk factors may have also contributed to the unequal variance between ages. I could not specifically test this here as, although it was stated that some subjects had hypertension in general descriptions of the ADNI and OASIS databanks, the actual measures of blood pressure and individual medical diagnoses for all subjects were not available to me. However, as at least 50% of subjects tend to be diagnosed with hypertension in many "normal" older cohorts (Aribisala et al., 2012), it could be argued that the inclusion of subjects with and without hypertension is more representative of the normal ageing population.

The incidence of normal subjects with silent AD pathology, e.g. that might be detected with Pittsburgh Compound-B (PIB) binding, or longitudinal cognitive decline that has not yet reached the point of dementia, may partially explain the increasing variance in brain structure with age. However, these data were not available for all subjects and therefore a bias would have been introduced had I excluded only some subjects based on these measures. Further, it was not the aim of this study to determine the source of brain structure variance in normal ageing but to demonstrate the effect of this variance. If the aim of this study had been to determine the source of variance then subjects with the measures could have been excluded and a sensitivity analysis done to detect bias. Future work with larger samples, more demographic and medical measures, and longitudinal data (to identify who eventually converts to disease) will be required to determine true population distributions and the sources of variance in normal ageing brain structure.

My calculation of TIV will have underestimated true TIV because the venous sinuses as well as CSF expand to occupy space vacated by the shrinking brain (Aribisala et al., 2012). Further, FMRIB's FAST may have incorrectly classified reduced signal areas of WM on MP-RAGE (potentially WM lesions) as GM. I did not have the MRI sequences, e.g. FLAIR, required to correct these errors for all subjects

(Marcus et al., 2007c). These regions were still correctly classified as tissue, i.e. not CSF, and so I combined GM and WM volumes to calculate whole brain tissue volume for each subject. Future studies and brain image databanks will need to acquire these additional sequences to define the true distributions of brain shrinkage and specific tissue, i.e. GM vs. WM, changes that occur with ageing.

Although all subjects were scanned with the same T1 MP-RAGE sequence, the different scanners used in ADNI may have also affected variance in brain volumes (Jovicich et al., 2009). Future work may implement scanner type as a random effect. However, all scanner protocols were standardised (Jack Jr et al., 2008) and the magnitude of difference between brains that I detected is too large to be attributed only to differences in scanner performance. Studies using one scanner and equally sized age groups (Farrell et al., 2009) have shown that irregular variance in brain tissue volume is attributable to advancing age.

Percentile rank models provide at least two new sources of information in brain image analyses. The first is that, given the general association between brain volume and cognitive function (Fotenos et al., 2005), percentile ranks may provide a measure to predict future brain loss and cognitive decline in individual patients, i.e. subjects at the lowest percentile ranks may be at greater risk of developing cognitive decline and dementia. I will test this in a planned longitudinal study. The second benefit is that percentile ranks may provide a deeper understanding of the differences between normal and diseased groups. For example, they will show whether general (mean) differences are due to consistent differences between subject groups or whether a given group of subjects has a skewed distribution of brain structure, e.g. a proportion of subjects with extremely low values of brain volume. These potential benefits, as well as the data I have presented here, suggest that percentile rank models may provide a deeper understanding of brain volume changes in normal ageing and, in future, assist diagnoses of neurodegenerative disease.

As neurological diseases are associated with subtle changes that may occur earlier in life (Shenton et al., 2001; Tondelli et al., 2012), I also assessed whether mean differences in regional brain volumes between younger adults were representative of percentile rank differences. This is described in the next chapter.

7. The accuracy of Gaussian distributions of regional brain volumes across early to middle adult life[5]

7.1 Introduction

This work (among that of many others, e.g. Manolio et al., 1994; Good et al., 2001; Ge et al., 2002; Resnick et al., 2003; Sowell et al., 2003; Walhovd et al., 2005a; Kruggel, 2006; Farrell et al., 2009) has demonstrated that the human brain changes with age and disease. As I showed in the previous chapter, brain structure does not change uniformly between individuals. Moreover, specific regions within the brain have differential rates of change (Walhovd et al., 2005b; Raz et al., 2010). For example, some people retain their youthful brain size into old age while others lose brain tissue much sooner (Farrell et al., 2009). Within subjects the effect size of age has been shown to be smaller in superior regions, e.g. the superior motor cortex, relative to inferior regions such as the medial temporal lobe, where the effect size of age on tissue volume was greater (Sowell et al., 2003; Allen et al., 2005).

As with whole brain volumes, differences in regional brain volumes between ages and states of health and disease were previously assessed with mean estimates and confidence intervals based on the Gaussian distribution (Allen et al., 2005; Raz et al., 2010). These parametric models assume that the distributions of brain volumes do not change with ROI, age, or disease (Meehl, 1978; Cohen, 1994; Freedman et al., 2007). Several independent studies have shown that the distributions of brain tissue do change with ROI location, age, and disease (Manolio et al., 1994; Ge et al., 2002; Resnick et al., 2003; Sowell et al., 2003; Thompson et al., 2003; Kruggel, 2006; Farrell et al., 2009; Raz et al., 2010). This suggests that mean differences may not reflect differences between subjects that are far from the mean.

Brain structure data are sometimes transformed, e.g. via "Box–Cox" (Box and Cox, 1964) or "nonparametric smoothing splines", to make them less vulnerable to deviations from model assumptions (Kruggel, 2006; Ziegler et al., 2011). But this

[5] Some of the work in this chapter was submitted for presentation as follows:

Dickie, D.A., Job, D.E., Wardlaw, J.M., Laidlaw, D.H., Bastin, M.E. (2014). Evidence of non-normal distributions in brain imaging data from normal subjects: implications for diagnosis of disease. Proc. Intl. Soc. Mag. Reson. Med. 22.

introduces an unnecessary layer of complexity (Freedman, 2010); may lead to irreproducible results depending on the transformation method used (Scott and Wild, 1991); and makes data difficult to interpret.

In other parametric models of regional brain volumes, the distribution of data and their proximity to the Gaussian distribution were not reported (Allen et al., 2005; Li et al., 2012). The robustness and generalisability of parametric models can vary dramatically when their assumptions are not met (Meehl, 1978; Cohen, 1994; Freedman et al., 2007). However, although applied often to model brain structure (Ashburner and Friston, 2000; Good et al., 2001), the effect of departures from parametric assumptions has not been thoroughly tested in these models.

Specifically, the differences between models that assume Gaussian distributions and equal variances (i.e. parametric models) and models that do not make these assumptions (i.e. certain nonparametric models) are unknown. Therefore, in this chapter I will compare parametric and nonparametric models of whole brain and regional volumes across early to middle adulthood (25–64 years).

7.2 Methods

The data described in this section were collected under NIH grant R01 EB004155-03 (PI: Dr. Mark E. Bastin, Neuroimaging Sciences, The University of Edinburgh).

7.2.1 Subjects

Eighty normal subjects (40 males, 40 females) aged 25–64 (median 43, IQR 17) years were recruited from advertisements in the Western General Hospital and Royal Infirmary, Edinburgh. All subjects gave written informed consent to be assessed and were determined to be normal via medical histories and a battery of cognitive tests. Their demographics are detailed in Table 21.

Table 21. Demographics and age group labels of young to middle adult sample

Age group in years	Age group label[1]	Number
25–34	1	21
35–44	2	23
45–54	3	24
55–64	4	12
25–64		80

Note: [1]this was the label used in statistical analyses described in the sections 7.2.5 and 7.2.6.

7.2.2 Cognitive tests

Each subject completed a battery of cognitive tests that were taken from the Wechsler Adult Intelligence Scale III (WAIS-III; Wechsler, 1997a), Wechsler Memory Scale III (WMS-III; Wechsler, 1997b), National Adult Reading Test (NART; Nelson and Willison, 1991), verbal fluency (Lezak et al., 2004), and simple and four–choice reaction time (Deary et al., 2001) tasks.

7.2.3 Brain MRI acquisition

Brain MRI data were acquired with a GE Signa Horizon HDxt 1.5T clinical scanner (General Electric, Milwaukee, WI, USA). The imaging protocol consisted of: axial T2–, T2*–, and FLAIR–weighted sequences, a coronal T1–weighted volume sequence, an axial T1–weighted fast–spoiled gradient echo (FSPGR) sequence, a magnetization transfer (MT–MRI) pulse sequence, and a diffusion MRI protocol, all acquired in Digital Imaging and Communications in Medicine (DICOM). Using MRIcron (Rorden, 2010), I converted the T1–weighted volume from DICOM to NIFTI–1 format (http://nifti.nimh.nih.gov/nifti–1) so to measure regional brain volumes.

7.2.4 Brain MRI processing

Regional brain volumes were extracted from each subject in native space and normalised by total intracranial volume (TIV) using the procedure described in chapter 5. Briefly, non–brain structure was removed by diffeomorphically warping (Avants et al., 2008) and applying the MNI152 brain atlas to each subject before manually correcting errors, e.g. remaining skull, slice–by–slice. Whole brain tissue

volumes were then calculated using FAST (Zhang et al., 2001). The SRI24 labelled brain atlas was then diffeomorphically warped to each subject and left and right amygdala, hippocampus, parahippocampal gyrus, caudate, putamen, and thalamus volumes were extracted from the whole brain GM volume. Because these structures represent a very small proportion of the TIV, sometimes much less than 1%, and to be consistent in tables throughout this chapter, I used scientific notation to present these results.

7.2.5 Parametric and nonparametric distributions of brain volumes

Parametric (Gaussian) distributions were calculated with the mean and standard deviation of brain volumes for each age group. Truly Gaussian distributions may be summarised as in Table 22 and Figure 3 in chapter 1.

Table 22. Summary of the Gaussian distribution

Percent of data (subjects)	Area of distribution
2.5–16	>= –2σ to –1σ
>16–84	> –1σ to +1σ
>84–97.5	> +1σ to +2σ

Note: σ=standard deviation

In other words, when data are Gaussian, the 2.5th percentile rank value is approximately equal to the mean minus 2 SD, the 16th percentile to the mean minus one SD, the 84th percentile to the mean plus one SD, and the 97.5th percentile to the mean plus 2 SD (Freedman et al., 2007). I used ±2 SD because ±1.96 SD sometimes underestimated the 95% limits (97.5th–2.5th percentile rank) of simulated Gaussian brain structure data (±2 SD was a slightly closer approximation).

I calculated percentile rank values using the ranking (nonparametric) approach described by equation 6,

$$np = j + g \tag{6.1}$$

$$y = 1/2(x_j + x_{j+1}) \quad \text{if } g = 0 \tag{6.2}$$

$$y = x_{j+1} \quad \text{if } g > 0$$

where n is the number of subjects, for the tth percentile $p=t/100$, j is the integer part of np, g is the fractional part of np, y is the tth percentile, and x_1, x_2, ... , x_n are the ordered values of each brain volume.

Parametric and nonparametric distributions of brain volumes were compared between age groups. That is, the parametric percentiles of each group were subtracted and the nonparametric percentiles were subtracted. These comparisons are described by equation 7,

$$1_{m,p}–2_{m,p} \tag{7.1}$$
$$1_{m,p}–3_{m,p} \tag{7.2}$$
$$1_{m,p}–4_{m,p} \tag{7.3}$$
$$2_{m,p}–3_{m,p} \tag{7.4}$$
$$2_{m,p}–4_{m,p} \tag{7.5}$$
$$3_{m,p}–4_{m,p} \tag{7.6}$$

where 1 is age group, 25–34 years, 2 is age group 35–44 years, 3 is age group 45–54 years, 4 is age group 55–64 years, m is method (parametric or nonparametric), and p is percentile rank, e.g. 2.5th; both m and p are the same within each comparison. In other words, I compared the percentile rank subtraction results between methods but not the differences between percentiles, e.g. 2.5th versus 97.5th.

Differences in parametric and nonparametric comparisons were computed using absolute percent error (equation 8)

$$\frac{\left|G_{c,p} - R_{c,p}\right|}{\left|G_{c,p}\right|} \times 100 \tag{8}$$

where G is the result from comparison c (e.g. age group 1–2) of the parametrically (Gaussian) defined percentile p (e.g. 2.5th), and R is the corresponding result from comparison c (e.g. age group 1–2) of the nonparametrically (ranking) defined percentile p. A sign was added to the percent error if the direction of the

comparison result differed between methods, i.e. if the parametric method said age group 1 had a larger value while the nonparametric method said age group 2 had a larger value.

7.2.6 Parametric and nonparametric effect sizes

The most commonly used parametric effect size is "*Cohen's d*", calculated by equation 9,

$$d = \frac{\mu_i - \mu_j}{\sigma_p} \tag{9}$$

where μ_i is the mean of group *i*, e.g. young adults, μ_j is the mean of group *j*, e.g. middle–aged adults, and σ_p is the pooled standard deviation of groups. Although a loose definition, parametric effect sizes of ≤ 0.2 are "small", 0.5 is described as "medium", and 0.8 is large (Cohen, 1988; Coe, 2002). I have extrapolated this definition to that in Table 23.

Table 23. Definition of effect sizes

Effect size	Definition
$d \leq 0.2$	Small
$0.2 < d < 0.5$	Small, medium
$d = 0.5$	Medium
$0.5 < d < 0.8$	Medium, large
$d \geq 0.8$	Large

For "*Cohen's d*" to be reliable it requires the data be Gaussian distributed (Cohen, 1988; Coe, 2002).

If data were Gaussian distributed then the result from equation 9 would approximate the result from the equivalent nonparametric equation (Freedman et al., 2007), equation 10

$$\tilde{d} = \frac{\tilde{\mu}_i - \tilde{\mu}_j}{\tilde{\sigma}_p} \tag{10}$$

where $\tilde{\mu}_i$ is the median of group *i*, $\tilde{\mu}_j$ is the median of group *j*, and $\tilde{\sigma}_p$ is the pool of the 50th percentile minus the 16th percentile in each group (when data are distributed Gaussian, the 50th percentile minus the 16th percentile is approximately equal to the standard deviation; Cohen, 1988; Freedman et al., 2007; Fritz et al., 2012).

Differences in parametric and nonparametric effect sizes were computed using absolute percent error (equation 11),

$$\frac{\left|d_{i,j} - \tilde{d}_{i,j}\right|}{\left|d_{i,j}\right|} \times 100 \tag{11}$$

where *d* is the parametrically defined effect size between groups *i* and *j*, and $\tilde{d}$ is the nonparametrically defined effect size between groups *i* and *j*. A sign was added to percent error if the direction of the effect differed between methods.

For consistency of reporting, and because of the potential for small ratios of brain volumes, four significant figures are generally reported, e.g., 0.0005 may round up to 0.001 but that rounding doubles the apparent size of the effect. When ratios are smaller still than four significant figures, scientific notation is used to report results.

7.3 Results

7.3.1 Standard deviations in whole and regional brain volumes

Standard deviations of whole brain volumes in each age group are listed in Table 24.

Table 24. Standard deviations of whole brain volumes in each age group

Age group	N	GM	%Err	CSF	%Err	WM	%Err
1	21	1.5E–02	13	1.3E–02	14	7.3E–03	–36
2	23	1.2E–02	–14	1.3E–02	12	9.2E–03	–20
3	24	1.5E–02	12	1.0E–02	–10	1.6E–02	36
4	12	9.3E–03	–31	6.9E–03	–40	1.1E–02	–1
Pooled	80	1.3E–02		1.1E–02		1.1E–02	

Note: N=number of subjects; GM=grey matter; WM=white matter; CSF=cerebrospinal fluid; %Err=percent error.

In general, there were no major difference between each age group and the pooled standard deviations. However, there were major differences between age groups, e.g. WM SD in age group 1 (25–34 years) was 36% smaller than the pooled WM SD whereas WM SD in age group 3 (45–54 years) was 36% larger, equating to a difference between these groups of 72% of the expected (pooled SD) value. This pattern of SD differences was repeated throughout the regional brain volumes, as shown in Table 25 and Table 26 (on the next page). The parametrically and nonparametrically derived distributions of these volumes are illustrated in section 7.3.2 and the quantitative comparison is in Appendix AI.3.

Table 25. Standard deviations of hippocampal complex volumes in each age group

Age	N	Hippo_L	%Err	Hippo_R	%Err	PHippo_L	%Err	PHippo_R	%Err	Amygdala_L	%Err	Amygdala_R	%Err
1	21	1.1E–04	3	1.4E–04	16	1.1E–04	–8	1.2E–04	2	4.0E–05	12	4.5E–05	7
2	23	1.2E–04	15	1.1E–04	–4	1.4E–04	17	1.2E–04	1	3.7E–05	3	4.4E–05	4
3	24	7.4E–05	–28	9.9E–05	–15	1.0E–04	–15	1.1E–04	–7	2.9E–05	–18	3.5E–05	–17
4	12	1.2E–04	14	1.2E–04	6	1.2E–04	4	1.3E–04	9	3.4E–05	–4	4.7E–05	13
Pooled	80	1.0E–04		1.2E–04		1.2E–04		1.2E–04		3.6E–05		4.2E–05	

Note: N=number of subjects; L=left; R=right; Hippo=hippocampus; PHippo=parahippocampal gyrus; ; %Err=percent error.

Table 26. Standard deviations of thalamus and basal ganglia volumes in each age group

Age	N	Caud_L	%Err	Caud_R	%Err	Puta_L	%Err	Puta_R	%Err	Thalamus_L	%Err	Thalamus_R	%Err
1	21	1.2E–04	5	1.8E–04	14	1.7E–04	8	2.0E–04	7	1.2E–04	–35	1.2E–04	–30
2	23	1.2E–04	4	1.4E–04	–11	1.3E–04	–18	1.7E–04	–7	2.2E–04	19	1.9E–04	15
3	24	1.1E–04	–1	1.7E–04	7	1.6E–04	2	1.9E–04	4	1.7E–04	–9	1.9E–04	14
4	12	9.7E–05	–16	1.2E–04	–24	1.7E–04	13	1.7E–04	–9	2.4E–04	27	1.3E–04	–19
Pooled	80	1.2E–04		1.5E–04		1.6E–04		1.8E–04		1.9E–04		1.7E–04	

Note: N=number of subjects; Caud=caudate; puta=putamen; L=left; R=right; %Err=percent error.

7.3.2 Parametric and nonparametric distributions of brain volumes

The parametric and nonparametric distributions of whole brain volumes are shown in Figure 34. It can be seen that the Gaussian distribution (left panel) is often not a good approximation for the shape of the actual data (right panel). Taken in conjunction with Table 24, these figures show that equality of variance (equal SDs) does not necessarily mean that the overall shape of data is the same between groups. For example, the SD of CSF volume was almost exactly the same between age group 1 and age group 3, however, age group 1 had a heavy left tail, high peak, and light right tail whereas age group had a flatter distribution and heavy right tail (top right panel, Figure 34).

This pattern is repeated throughout the regional brain volumes. For example, the SD of right parahippocampal volume was almost exactly the same between age group 1 and age group 2, and the shapes of their distributions were somewhat similar (bottom right panel, Figure 36). However, the SD of right thalamus was almost exactly the same between age group 2 and age group 3, whereas the shapes of their distributions were markedly different (bottom right panel, Figure 38). This further illustrates that, while it may in some instances, equality of variance (equal SDs) does not necessarily mean that the overall shape of data is the same between groups.

The differences in distribution shape in the following figures (only figures and no main body text appear on the next two pages) show that approximating brain volume data with the Gaussian distribution and equal variances sometimes leads to the smoothing away of differences between groups, and other times creates differences that do not exist in the actual data. This is quantitatively demonstrated in Appendix AI.3.

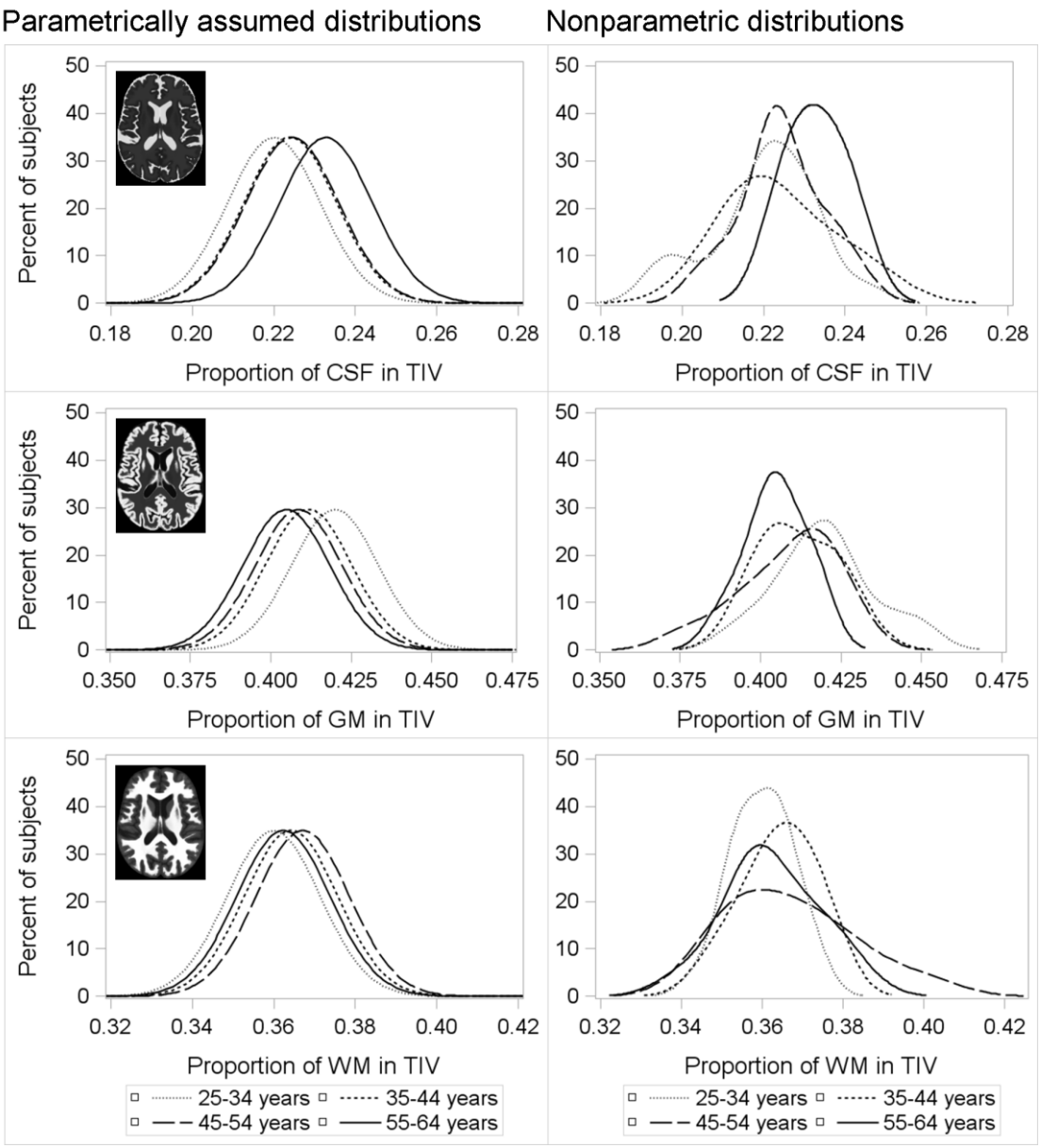


Figure 34. Parametric and nonparametric distributions of whole brain volumes in each age group
GM=grey matter; WM=white matter; CSF=cerebrospinal fluid; TIV=total intracranial volume; 26–35 years=age group 1; 36–45 years=age group 2; 46–55 years==age group 3; 56–65 years=age group 4.

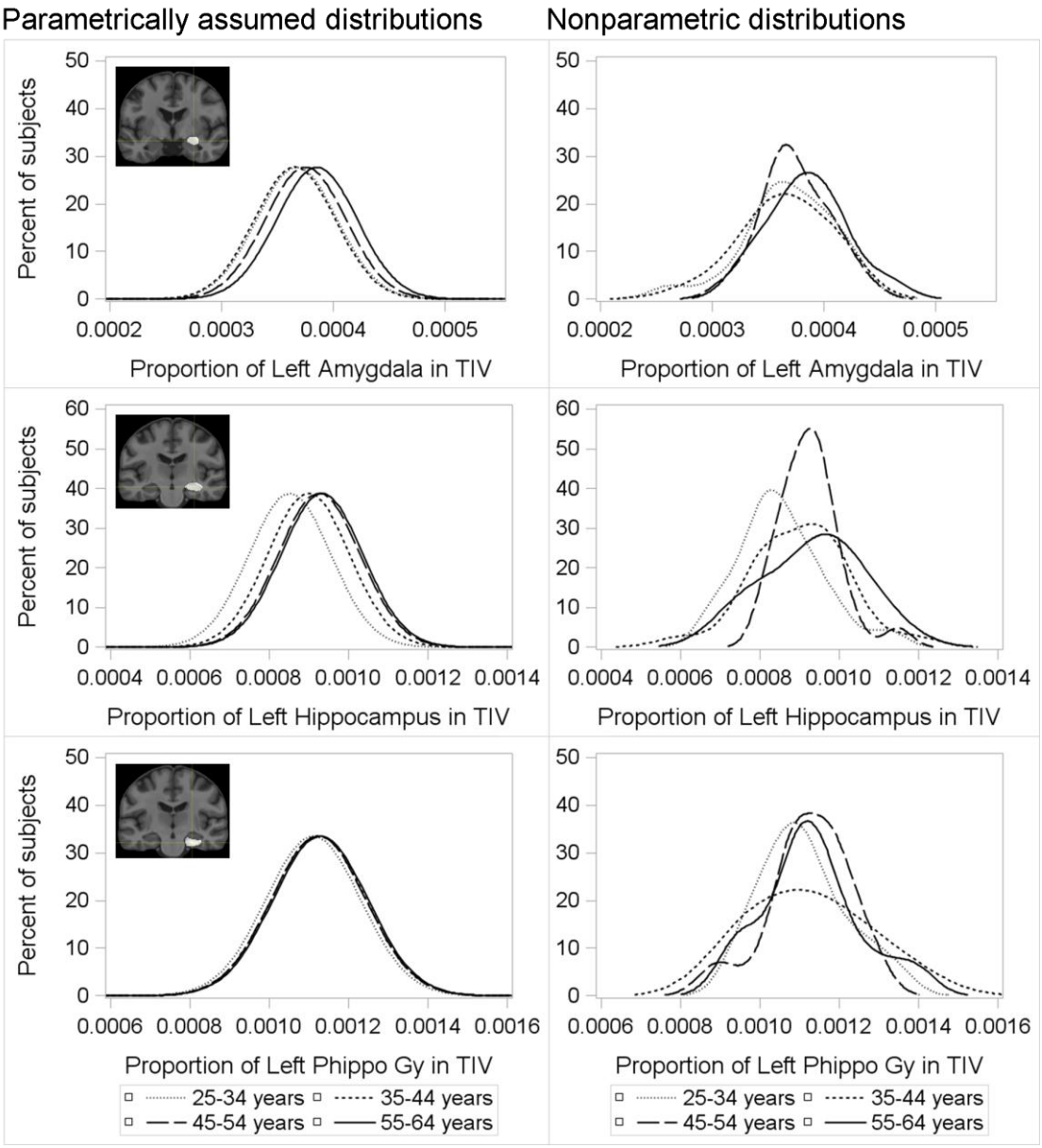


Figure 35. Parametric and nonparametric distributions of left hippocampal complex volumes in each age group

Phippo Gy=parahippocampal gyrus; TIV=total intracranial volume; 26–35 years=age group 1; 36–45 years=age group 2; 46–55 years==age group 3; 56–65 years=age group 4. The brain images are shown in radiological convention, i.e. left on the image is the patient's right.

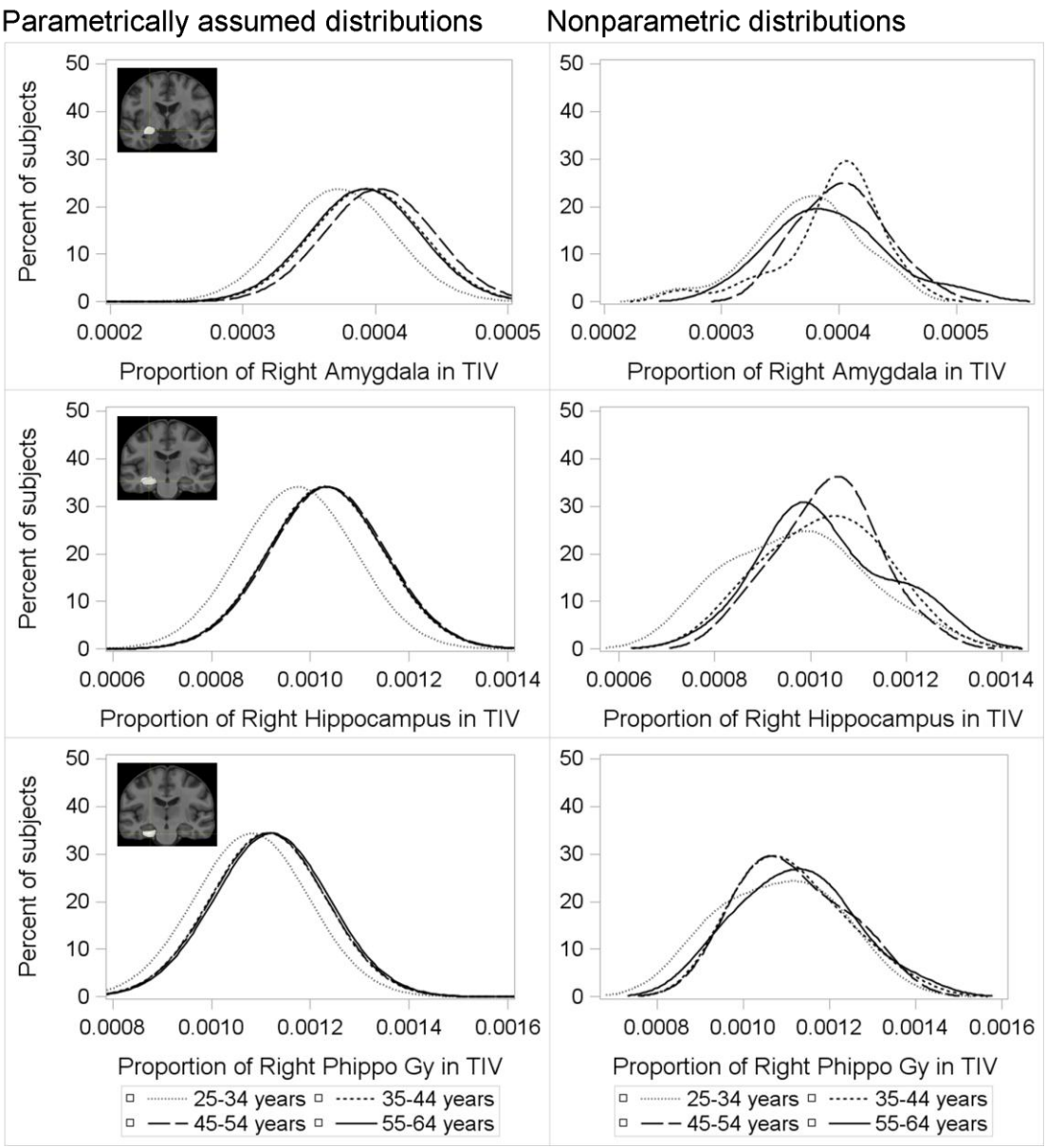


Figure 36. Parametric and nonparametric distributions of right hippocampal complex volumes in each age group

Phippo Gy=parahippocampal gyrus; TIV=total intracranial volume; 26–35 years=age group 1; 36–45 years=age group 2; 46–55 years==age group 3; 56–65 years=age group 4. The brain images are shown in radiological convention, i.e. left on the image is the patient's right.

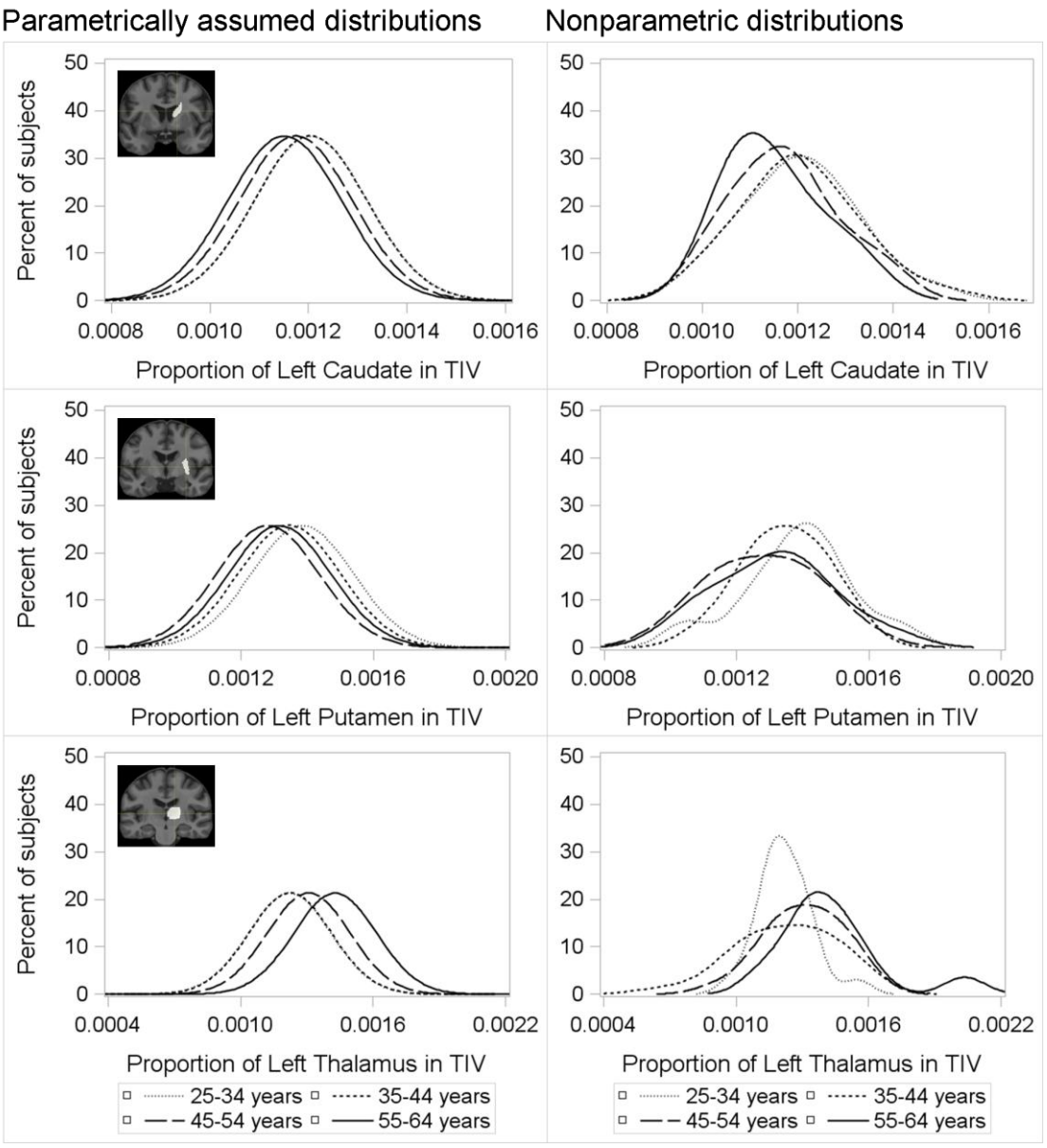


Figure 37. Parametric and nonparametric distributions of left thalamus and basal ganglia volumes in each age group

26–35 years=age group 1; 36–45 years=age group 2; 46–55 years==age group 3; 56–65 years=age group 4. The parametric distributions of the left thalamus in subjects 25-34 and 35-44 years are almost entirely overlapping (bottom left). The brain images are shown in radiological convention, i.e. left on the image is the patient's right.

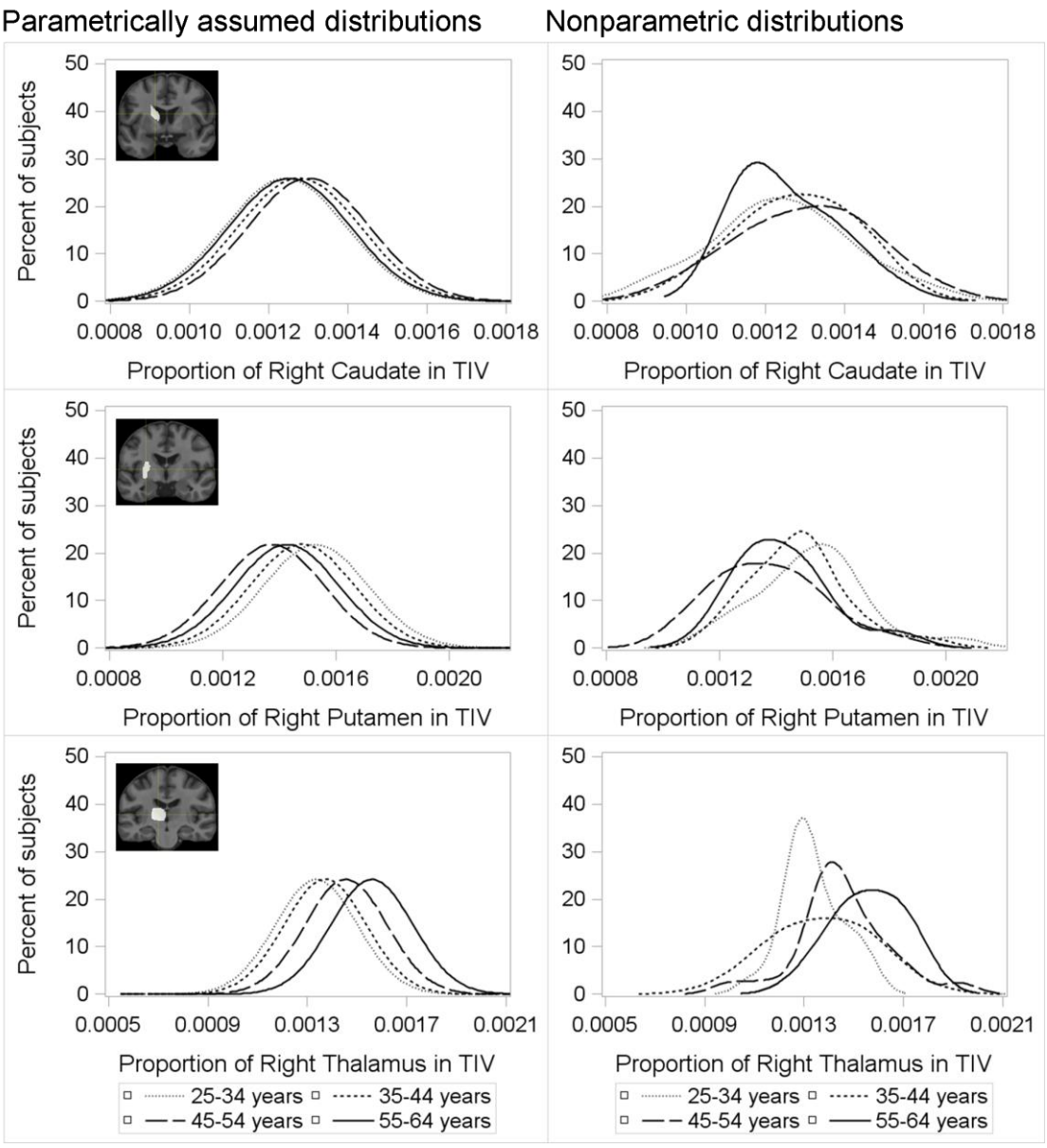


Figure 38. Parametric and nonparametric distributions of right thalamus and basal ganglia volumes in each age group
26–35 years=age group 1; 36–45 years=age group 2; 46–55 years==age group 3; 56–65 years=age group 4. The brain images are shown in radiological convention, i.e. left on the image is the patient's right.

7.3.3 Comparisons between groups according to parametric and nonparametric distributions

The effect of differences in SD and overall distribution shape is quantitatively demonstrated in this section. The percent error column (%Err) in the tables in Appendix AI.3 indicates the absolute percentage differences between parametric and nonparametric percentile comparisons. Table AI.3.1.1 lists whole brain volume distributions, Tables AI.3.2 and AI.3.3 list hippocampal complex distributions, and Tables AI.3.4 and AI.3.5 list basal ganglia and thalamus distributions. Negative percent errors indicate that the

parametric and nonparametric methods differed on the direction of the difference between groups. For example, in Table AI.3.1.1, the mean difference in CSF between groups 1 and 2 was 139% larger than the median difference and was in the opposite direction, i.e. the mean difference said CSF volume was greater in group 2, whereas the median difference said CSF was less in group 2.

Overall these tables demonstrate that the limits, i.e. 2.5th percentiles, of whole and regional brain volumes are often not well approximated by parametric methods, e.g. by extrapolating limits from the mean and standard deviation (see section 7.2.5 for a detailed description of how this is done). Further, there was no systematic difference between the methods: sometimes the parametric method reduced differences in the actual data while other times it exaggerated differences in the actual data.

7.3.4 Parametric and nonparametric effect sizes

The percent error column (%Err) in the following three tables indicates the absolute differences between parametric and nonparametric effect sizes. Negative percent errors indicate that the parametric and nonparametric methods differed on the direction of the effect between groups.

The parametric and nonparametric effect sizes in whole brain volumes between age groups are shown in Table 27. Although some of these percent errors are very large, the general classification of effect (i.e. “small”, “medium”, or “large”; Cohen, 1988) between the methods is the same, e.g. the parametric effect of –0.027 in GM between groups 2 and 3 and the nonparametric effect of –0.175 are both considered “small”. Therefore, I highlighted in bold entries in tables 34, 35, and 36 where the differences in effect size between methods transcended classification boundaries. As discussed in the methods, effect sizes of ≤0.2 are “small”, 0.5 is described as “medium”, and 0.8 is large (see section 7.2.6 in the methods for further details; Cohen, 1988; Coe, 2002). This illustrates that of all tissue volume comparisons, parametric and nonparametric effect sizes in grey matter were often particularly different.

Table 27. Parametric and nonparametric effect sizes in whole brain volumes

Volume	Comparison	Parametric	Nonparametric	%Err
CSF	**"1–2"**	**–3.44E–01**	**1.30E–01**	**–138**
	"1–3"	**–3.71E–01**	**–4.58E–02**	**88**
	"1–4"	–1.11E+00	–7.82E–01	30
	"2–3"	–2.73E–02	–1.75E–01	543
	"2–4"	–7.68E–01	–9.12E–01	19
	"3–4"	–7.41E–01	–7.37E–01	1
GM	**"1–2"**	**6.09E–01**	**9.20E–01**	**51**
	"1–3"	**8.43E–01**	**5.43E–01**	**36**
	"1–4"	1.12E+00	1.12E+00	0
	"2–3"	2.34E–01	–3.77E–01	–261
	"2–4"	**5.12E–01**	**2.00E–01**	**61**
	"3–4"	**2.77E–01**	**5.77E–01**	**108**
WM	"1–2"	–3.75E–01	–4.14E–01	11
	"1–3"	**–6.23E–01**	**–4.30E–01**	**31**
	"1–4"	–2.10E–01	5.09E–02	–124
	"2–3"	–2.49E–01	–1.52E–02	94
	"2–4"	**1.65E–01**	**4.65E–01**	**183**
	"3–4"	4.13E–01	4.81E–01	16

Note: GM=grey matter; WM=white matter; CSF=cerebrospinal fluid; %Err=percent error.

The parametric and nonparametric effect sizes in hippocampal complex volumes between age groups are shown in Table 28. Effect sizes in the left hippocampus and parahippocampal gyrus, and left and right amygdala were often reasonably similar between parametric and nonparametric methods. However, effect sizes in the right hippocampus and parahippocampal gyrus were often different between methods.

Table 28. Parametric and nonparametric effect sizes in hippocampal complex volumes

Volume	Comparison	Parametric	Nonparametric	%Err
Hippocampus_L	**"1–2"**	**–4.57E–01**	**–7.40E–01**	**62**
	"1–3"	–7.09E–01	–7.60E–01	7
	"1–4"	–7.86E–01	–9.67E–01	23
	"2–3"	**–2.52E–01**	**–1.99E–02**	**92**
	"2–4"	–3.29E–01	–2.27E–01	31
	"3–4"	–7.73E–02	–2.07E–01	167
ParaHippocampal_L	"1–2"	–1.08E–01	–2.01E–01	86
	"1–3"	**–1.08E–01**	**–2.54E–01**	**135**
	"1–4"	–1.57E–01	–1.23E–01	22
	"2–3"	–8.40E–04	–5.37E–02	6294
	"2–4"	–4.96E–02	7.75E–02	–256
	"3–4"	–4.87E–02	1.31E–01	–369
Amygdala_L	"1–2"	5.27E–02	–2.36E–02	–145
	"1–3"	–2.11E–01	–1.99E–01	5
	"1–4"	–4.94E–01	–5.96E–01	21
	"2–3"	**–2.63E–01**	**–1.76E–01**	**33**
	"2–4"	–5.46E–01	–5.72E–01	5
	"3–4"	–2.83E–01	–3.96E–01	40
Amygdala_R	"1–2"	–5.16E–01	–5.83E–01	13
	"1–3"	–7.28E–01	–5.10E–01	30
	"1–4"	–4.71E–01	–2.08E–01	56
	"2–3"	–2.11E–01	7.37E–02	–135
	"2–4"	**4.57E–02**	**3.75E–01**	**722**
	"3–4"	2.57E–01	3.02E–01	17
Hippocampus_R	"1–2"	–4.78E–01	–4.74E–01	1
	"1–3"	**–5.29E–01**	**–4.41E–01**	**17**
	"1–4"	**–4.99E–01**	**–9.14E–02**	**82**
	"2–3"	–5.04E–02	3.25E–02	–164
	"2–4"	**–2.05E–02**	**3.82E–01**	**–1964**
	"3–4"	**2.99E–02**	**3.50E–01**	**1069**
ParaHippocampal_R	**"1–2"**	**–2.79E–01**	**3.73E–04**	**–100**
	"1–3"	**–2.95E–01**	**–5.38E–02**	**82**
	"1–4"	**–3.52E–01**	**–1.34E–01**	**62**
	"2–3"	–1.55E–02	–5.42E–02	249
	"2–4"	–7.24E–02	–1.34E–01	85
	"3–4"	–5.69E–02	–8.00E–02	41

Note: ParaHippocampal=parahippocampal gyrus.

The parametric and nonparametric effect sizes in thalamus and basal ganglia volumes between age groups are shown in Table 29.

Table 29. Parametric and nonparametric effect sizes in thalamus and basal ganglia volumes

Volume	Comparison	Parametric	Nonparametric	%Err
Caudate_L	"1–2"	7.83E–03	8.75E–02	1018
	"1–3"	2.72E–01	2.79E–01	2
	"1–4"	**4.64E–01**	**7.16E–01**	**54**
	"2–3"	2.64E–01	1.91E–01	28
	"2–4"	**4.57E–01**	**6.28E–01**	**38**
	"3–4"	**1.92E–01**	**4.37E–01**	**127**
Putamen_L	"1–2"	2.59E–01	3.15E–01	22
	"1–3"	6.81E–01	6.64E–01	2
	"1–4"	4.46E–01	3.81E–01	15
	"2–3"	4.22E–01	3.49E–01	17
	"2–4"	1.87E–01	6.59E–02	65
	"3–4"	–2.35E–01	–2.83E–01	21
Thalamus_L	"1–2"	–1.23E–02	–2.05E–01	1566
	"1–3"	**–4.57E–01**	**–6.18E–01**	**35**
	"1–4"	–1.10E+00	–1.00E+00	9
	"2–3"	–4.44E–01	–4.13E–01	7
	"2–4"	–1.09E+00	–7.98E–01	27
	"3–4"	**–6.42E–01**	**–3.85E–01**	**40**
Caudate_R	"1–2"	–2.22E–01	–3.44E–01	55
	"1–3"	**–3.75E–01**	**–5.32E–01**	**42**
	"1–4"	**–6.82E–02**	**5.50E–01**	**–906**
	"2–3"	–1.53E–01	–1.88E–01	23
	"2–4"	**1.54E–01**	**8.94E–01**	**481**
	"3–4"	**3.07E–01**	**1.08E+00**	**252**
Putamen_R	"1–2"	2.66E–01	3.21E–01	21
	"1–3"	8.54E–01	1.02E+00	19
	"1–4"	5.56E–01	7.96E–01	43
	"2–3"	5.88E–01	6.96E–01	18
	"2–4"	2.90E–01	4.74E–01	64
	"3–4"	–2.98E–01	–2.22E–01	26
Thalamus_R	**"1–2"**	**–2.15E–01**	**–5.19E–01**	**141**
	"1–3"	**–7.15E–01**	**–1.02E+00**	**42**
	"1–4"	–1.36E+00	–1.91E+00	41
	"2–3"	–4.99E–01	–4.97E–01	1
	"2–4"	–1.14E+00	–1.39E+00	22
	"3–4"	**–6.41E–01**	**–8.96E–01**	**40**

Effect sizes in the left and right putamen were generally similar between parametric and nonparametric methods. Although this was also true for the left thalamus, there were greater discrepancies between methods in the right thalamus, and right and left caudate.

7.4 Discussion

I have shown that parametric and nonparametric distributions of subregional brain volumes can be different, i.e. not only is the effect of age different across the brain, these regions do not have a uniformly Gaussian distribution. This suggests that parametric statistical models may not adequately represent the limits of brain volumes across adulthood. Some of the differences I found were exceptionally large and this is likely in part due to the limited number of subjects assessed. However, as demonstrated in the review in chapter 4, this number of subjects is not unusual for structural brain imaging analyses. Although these data may be transformed to better fit parametric models (Kruggel, 2006; Ziegler et al., 2011), the consistent differences I found suggest that these transforms may obscure the true limits of regional brain volumes.

Further to implications for defining the limits of normal brain volumes, and as I demonstrated by modelling effect sizes between ages, these changing and non-Gaussian distributions also have implications for central tendency analysis. Although the focus of this chapter was regional volumes, the parametric method led to consistent and large overestimation of the effect of age in whole brain GM. Within the hippocampal complex, the parametric method produced consistent overestimation of the effect of age in the right hippocampus and parahippocampal gyrus. This is especially noteworthy given that the hippocampus is often implicated in diagnoses of neurological diseases such as AD and schizophrenia (Jack Jr et al., 1997; Job et al., 2002).

However, not all regions exhibited great differences between methods. In the basal ganglia, and in particular the putamen, and also in the amygdala, parametric and nonparametric effect sizes generally agreed. While these central tendency measures agreed, the parametric and nonparametric percentile ranks of these regions did not agree. This (further to the data presented in chapter 6) illustrates that differences between subjects that are removed from the mean are not well represented by parametric models. It is therefore apparent that, to robustly define the range of effect sizes across the brain, nonparametric models may have to be used in conjunction with the more commonly applied parametric models.

By using the mean ± two SD, I may have slightly overestimated parametric limits. The associated error was 0.2% and this did not impact the large differences (>100%) I observed between parametric and nonparametric limits. I undertook this comparison in a single, relatively small sample. The oldest age group was smaller than the other groups so cannot be used for definitive measures but illustrates the potential dangers of making

these assumptions in similarly sized samples (that are common in brain imaging, chapter 4). The bimodal nature of the nonparametric distributions (e.g., in Figure 34) may have been due in part to the limited number of subjects (n ~20 per group). However, there is evidence from much larger studies (n>4000) of other biological variables (e.g., serum calcium; Elveback et al., 1970) and from other neuroimaging studies (Rorden et al., 2007) that suggest bimodal distributions of neuroimaging data are plausible. Future studies with much larger samples, e.g. at least 1 subject per percentile, will be required to confirm whether population distributions of neuroimaging data are actually bimodal.

The structural brain data I used were acquired using atlas based segmentation. Although I visually inspected results to ensure that there were no gross errors, there may have been more subtle errors that contributed to the unequal variance and lack of conformance to the Gaussian distribution. However, this is a commonly used method to extract regional brain volumes (Maldjian et al., 2003). I have illustrated that if data are acquired by this method then they may not be suitable for parametric statistical analyses. Future work will determine whether the method of segmentation contributes to the variance and distribution of regional brain volumes.

Although regional brain volumes may provide more sensitive markers for disease than whole brain volumes (Jack Jr et al., 1997; Fox and Schott, 2004), the differences between groups may be even more subtle than entire regional brain volumes, i.e. they may exist within regions at the voxel level (Job et al., 2002). I therefore developed a series of voxel-based statistical methods for assessing the subtlest of differences between normal and diseased individuals and groups.

8. Robust voxel–based statistics for identifying subtle brain pathology in individuals and groups[6]

8.1 Introduction

It is clear that human brain structure changes with age. But as I (and many others) have demonstrated, these changes are not uniform and their pattern and consequences are not fully understood (chapters 6 and 7; Ge et al., 2002; Resnick et al., 2003; Sowell et al., 2003; Allen et al., 2005; Farrell et al., 2009; Ferguson et al., 2010). A wide range of atrophy and other anomalies, e.g. white matter hypointensities/ hyperintensities on T1/ T2, can be seen in ageing brain magnetic resonance imaging (MRI). This is true even between subjects of the same age and with no apparent difference in cognitive or physical function (Farrell et al., 2009; Aribisala et al., 2012). It can then be difficult to establish whether whole and regional brain volumes are normal for age or suggestive of neurodegenerative disease (Shenton et al., 2001; Fotenos et al., 2005; Farrell et al., 2009; Ferguson et al., 2010). In these cases, voxel–based assessment/ classification may be a more sensitive marker for disease (Job et al., 2002).

Voxel–based null hypothesis significance testing (VBNHST) was developed to determine differences in MR brain structure between normal and diseased groups, but has recently been used to compare individual subjects to a clinically normal dataset (Ashburner and Friston, 2000; Scarpazza et al., 2013). To be robust, parametric VBNHST requires that data meet a number of assumptions about their generation and distribution, e.g. that they were randomly generated, and are equally Gaussian between subject groups (homoscedastic; Elveback et al., 1970; Freedman et al., 2007). These assumptions may not always be met in group studies of MR brain image data (Nichols and Holmes, 2002; Rorden et al., 2007) and VBNHST may not be reliable in individual subject analyses, e.g. it leads to a number of false positive

[6] Some of the work described in this chapter was presented as follows:

Dickie, D.A., Job, D.E., Rodríguez González, D., Shenkin, S.D., Wardlaw, J.M. (2013). How normal is this brain? Development and testing of a new MR template for voxel–based brain ranking. *Proceedings of the 19th Annual Meeting of the Organization for Human Brain Mapping*. Seattle, USA.

results (Scarpazza et al., 2013). Further, VBNHST was not originally designed to be used for individual subject analyses (Ashburner and Friston, 2000).

Rather than comparing individual subjects to a clinically normal dataset with VBNHST, others have directly calculated the placement of individuals relative to the distribution of clinically normal values (Pernet et al., 2009; Aubert-Broche et al., 2011). These studies estimated the distributions of clinically normal voxel values for younger subjects (<60 years) based on the Gaussian distribution. However, as I have shown, brain structure data may not always follow the Gaussian distribution (especially at older ages).

Further to deviations from assumed distributions, significance values, e.g. *P*–values from *t*–tests, may be difficult to interpret if the sample of subjects was not derived randomly. In such cases it is useful to calculate effect sizes between groups and compare these between cohorts (Meehl, 1978; Cohen, 1994; Freedman et al., 2007). One of the most widely used measures of effect size is "*Cohen's d*"; this is essentially the mean difference between groups normalised by their pooled SD (Cohen, 1988; Coe, 2002). As it is based on the standardised mean difference, this statistic still relies on the data being equally Gaussian between groups. Even small departures from the Gaussian distribution may lead to uninterpretable and/or unreliable results (Cliff, 1993; Coe, 2002). In these cases, it may be more appropriate to calculate the so called, "*d* statistic" – a type of nonparametric effect size that calculates the overlap between the distributions of two groups (Siegel and Castellan, 1988; Cliff, 1993; Coe, 2002). Although the overlap statistic is referred to as "*d*" in the literature, I use the Greek equivalent "*delta*" (δ) to avoid confusion with "*Cohen's d*". Despite not providing a measure of central tendency shift, i.e. mean or median shift, *delta* is still a clinically relevant and useful measure in itself, given that central tendency measures may not adequately describe the differences between two clinical groups (Elveback et al., 1970).

In this chapter I assess parametrically and nonparametrically defined normal reference distributions (normal atlases) for determining "how normal" voxel values are in individual subjects. I also develop methods for assessing the distributions of voxel values so to determine their suitability for subsequent parametric or nonparametric analyses. Finally, I implement voxel–based measures of effect size (*Cohen's d* and *delta*) to quantify the differences between normal and diseased groups.

8.2 Methods

8.2.1 Subjects and brain MRI

Brain MRI from 236 normal subjects and 224 subjects diagnosed with AD (all aged between 55 and 90 years) were acquired from ADNI (AD=124; controls=138) and OASIS (AD=100; controls=98). The normal subjects did not have dementia, but potentially had non–debilitating conditions common in ageing, e.g. hypertension. Their demographics were more thoroughly described in section 5.2.3. At the time of the present analyses, ADNI and OASIS were the only public sources of structural MRI brain scans with clinical metadata that fully represented normal older people (chapter 2). Simulation and permutation testing may provide an indication of the representativeness of such small amounts of data (Freedman and Lane, 1983). Permutations of the normal ageing GM atlas are presented later (section 8.3.4.1).

Both ADNI and OASIS provided 1.5T MP–RAGE T1–weighted MR brain images that were acquired in the sagittal plane at approximately 1x1x1mm resolution. I removed non–brain structure from the images before extracting grey matter proportion images (as described in chapter 5). I then calculated the range of voxel–based statistics described in the following sections.

8.2.2 Voxel–based statistics

I developed methods to calculate a number of parametric and nonparametric voxel–based statistics of grey matter proportion but, given that parametric statistics rely on data being Gaussian distributed (Elveback et al., 1970; Freedman et al., 2007), I first developed a method to assess the shape of voxel–wise data distributions.

8.2.2.1 Assessing voxel–wise data distributions

The distribution of data can be assessed with statistical tests such as the Kolmogorov–Smirnov test (Massey Jr, 1951) but they do not thoroughly describe the shape of data. These tests mainly state that the population data may or may not be shaped like the Gaussian distribution, e.g. they do not overtly specify whether the sample data are sharply peaked, or their mode is skewed to the left. Further, it is highly questionable whether the sample sizes studied in this work are sufficient to make true population inferences (Cohen, 1994; Freedman et al., 2007). Until we have sufficient population

data and to ensure that correct statistical models are used in the interim, it is important to understand how the available sample data are distributed. Whether or not sample data are Gaussian distributed and the shape of their distribution can be determined by calculating kurtosis and skewness.

8.2.1.1.1 Kurtosis

Kurtosis illustrates the central and outer appearance of a data distribution (DeCarlo, 1997). It is calculated by equation 12,

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{\sigma^4} \tag{12}$$

where x_{i-n} are all the values of variable x, μ is the mean of variable x, and σ is the standard deviation of variable x. The Gaussian distribution has kurtosis of 3, distributions with kurtosis less than 3 are flatter with lighter tails, and distributions with kurtosis more than three have higher peaks and heavier tails (DeCarlo, 1997).

8.2.1.1.2 Skewness

Skewness measures the symmetry of a data distribution. It is calculated by equation 13,

$$\frac{\sum_{i=1}^{n}(x_i-\mu)^3}{\sigma^3} \tag{13}$$

where x_{i-n} are all the values of variable x, μ is the mean of variable x, and σ is the standard deviation of variable x.

When skewness is equal to zero the data are perfectly symmetrical (like the Gaussian, bimodal, or uniform distributions); when it is above zero the data have a heavy left tail; and when it is below zero the data have a heavy right tail (Figure 39). The shape of distributions, as determined by kurtosis and skewness, can inform the most appropriate method for summarising them, i.e. via a parametric or nonparametric method.

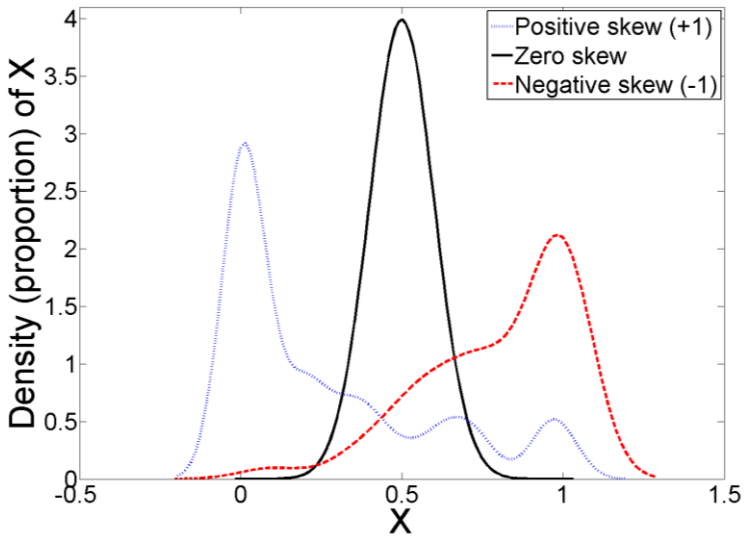


Figure 39. Examples of data (X) with +1 (blue dotted line), zero (black solid line), and –1 (dashed red line) skewness

8.2.2.2 Summarising voxel–wise data distributions for a normal reference brain atlas

Most previous brain atlases were constructed by calculating parametric (mean and standard deviation) voxel values. Parametric values are assumed to approximate the distribution of voxel values, i.e. equate to nonparametric percentile ranks. I tested this by constructing parametric and nonparametric distributional brain atlases of normal ageing brain structure to assess brain structure in older subjects with AD.

To derive the atlases I registered all atlas subjects (*n*=98 randomly selected normal ageing subjects, 49 ADNI; 49 OASIS) to the standard space described in chapter 5 via "nonlinear surface" (Nsurf) registration (also discussed in chapter 5). These normal ageing atlas subjects were not used for any other analyses in this chapter. Once all subjects were in this space I calculated parametric and nonparametric measures of voxel–wise distributions of grey matter proportion.

8.2.2.2.1 Parametric brain atlas

Parametric summaries of distributions are based on calculations of the data mean and standard deviations.

The percentile rank values of these distributions are approximated as in equation 14,

p		
2.5^{th}	$\mu - (2 \times \sigma)$	(14.1)
25^{th}	$\mu - (0.7 \times \sigma)$	(14.2)
50^{th}	μ	(14.3)
75^{th}	$\mu + (0.7 \times \sigma)$	(14.4)
97.5^{th}	$\mu + (2 \times \sigma)$	(14.5)

where p is the percentile rank, μ is the mean, and σ is the standard deviation. I used ±2 SD because ±1.96 SD sometimes underestimated the 95% limits (97.5^{th}–2.5^{th} percentile rank) of simulated Gaussian brain structure data. There was almost no discernable difference between atlases based on ±2 SD versus ±1.96 SD, but ±2 SD provided assessments slightly closer to the "true" nonparametric percentile ranks. For example, in simulated GM proportion data there were 1,847,047 voxels in agreement between parametric and nonparametric methods when using the ±1.96 SD atlas, and 1,847,686 voxels in agreement between methods when using the ±2 SD atlas.

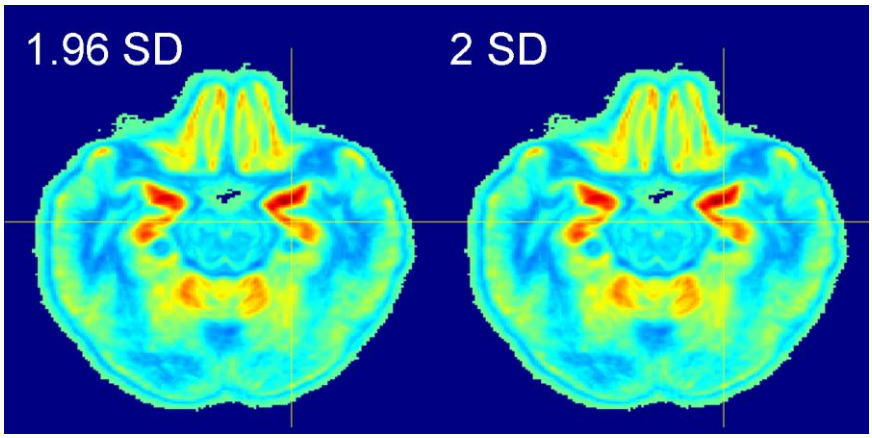


Figure 40. Parametric simulations of the 2.5th percentile rank of grey matter proportion using 1.96 SD (left panel) and 2 SD (right panel)

If data follow the Gaussian distribution, the parametrically estimated values of these percentile ranks should be approximately equal to the rank calculated nonparametric values (Elveback et al., 1970; Freedman et al., 2007; Freedman, 2010).

8.2.2.2.2 Nonparametric brain atlas

Nonparametric methods use a ranking approach to directly calculate percentile rank values as in equation 15,

$$np = j + g \tag{15.1}$$

$$y = 1/2(x_j + x_{j+1}) \quad \text{if } g = 0 \tag{15.2}$$

$$y = x_{j+1} \quad \text{if } g > 0$$

where n is the number of subjects, for the tth percentile $p=t/100$, j is the integer part of np, g is the fractional part of np, y is the tth percentile, and x_1, x_2, ... , x_n are the ordered values of each brain volume.

8.2.2.3 Assessing brain structure with a normal reference brain atlas: voxel classification

Much like clinical brain atlases are used to diagnose whole brain atrophy in patients (Farrell et al., 2009), I used voxel–based atlases to classify brain structure voxel–by–voxel in individual subjects. AD subjects (n=138, 89 ADNI; 49 OASIS) were registered to the atlas using Nsurf registration (described in chapter 5). In each voxel in atlas space, subject tissue proportion was compared to the tissue proportions of each level in the atlas and then assigned the rank to which their tissue proportion was nearest (Figure 41).

Note: although I illustrate this method using grey matter proportion images from T1-weighted images of normal older subjects and subjects diagnosed with AD, the software I developed can produce a percentile rank reference for any image sequence from any group, e.g. it can produce a T2–weighted percentile rank atlas from normal young adults for assessing brain structure in individuals and groups with schizophrenia or a T1–weighted percentile rank atlas from full term infants to assess brain structure in preterm individuals and groups.

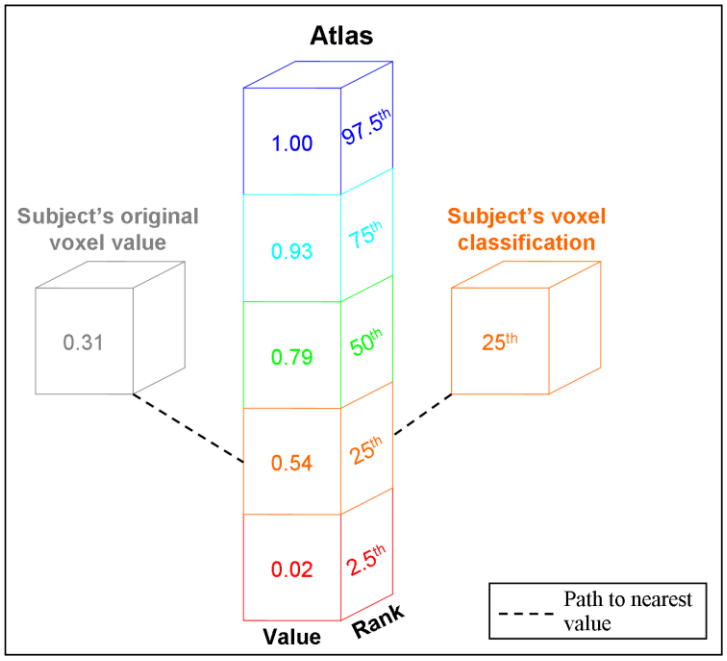


Figure 41. Schematic of voxel–based classification of brain structure with a normal reference atlas
The subject's original voxel value is compared to each of the atlas percentile rank values and is then assigned the rank to which its value is nearest.

Further to generating voxel–wise statistics for individual subject analyses, I developed methods to compute voxel–wise effect sizes for group studies.

8.2.2.4 Voxel–wise effect sizes for group studies

There are parametric and nonparametric methods for computing effect size in group studies. I sought to determine the difference in these measures in a study of GM proportion between normal ageing (*n*=138, 89 ADNI; 49 OASIS) and AD (*n*=138, 89 ADNI; 49 OASIS).

8.2.2.4.1 Parametric effect size: Cohen's d

When the random sample assumption of statistical tests is not met it is useful to calculate effect sizes between case and control groups so that results from replicated studies can be compared and validated (Meehl, 1978; Cohen, 1988; Siegel and Castellan, 1988; Cliff, 1993; Cohen, 1994; Coe, 2002; Freedman, 2010; Fritz et al., 2012).

The most commonly used parametric effect size, *Cohen's d*, is calculated by equation 16,

$$d = \frac{\mu_{cntrl} - \mu_{case}}{\sigma_p} \tag{16}$$

where μ_{cntrl} is the mean of the control group, μ_{case} is the mean of the case group, and σ_p is the pooled standard deviation of the two groups. Although a loose definition, parametric effect sizes of ≤ 0.2 are "small", 0.5 is described as "medium", and 0.8 is large (Cohen, 1988; Coe, 2002). If the data are Gaussian, a positive parametric effect means that, across the distribution, the control group is $d \times$SD (σ_p) larger than the case group, e.g. for an effect size of d=0.8 between controls and AD subjects, each percentile rank of the control group is 0.8 SD larger than the AD group.

If data are not equally Gaussian the parametric effect size becomes difficult to interpret (Coe, 2002), e.g. the standard difference between each group's percentiles (2.5^{th}, 25^{th}, …, 97.5^{th}) will likely not be equal to the (mean) effect size.

8.2.2.4.2 Nonparametric effect size: delta

When data are not equally Gaussian, the effect size, "*delta*", may be a more reliable indicator of the effect of a variable (note this is actually referred to as "*the d statistic*" but to avoid confusion with "*Cohen's d*", I use the Greek equivalent "*delta*", δ). Delta does not require data to be Gaussian or equally distributed (homoscedastic) between groups (Siegel, 1957; Cliff, 1993; Coe, 2002) and is calculated by equation 17

$$\delta = \frac{\#(x_i > x_j) - \#(x_i < x_j)}{mn} \tag{17}$$

where all values of x in group i are compared with all values of x in group j, and mn is the total number of comparisons (number of subjects in group i × number of subjects in group j). The interpretation of delta is that it is a direct measure of how often one group, e.g. a control group, has a higher value than another group, e.g. a group of AD subjects. For example, δ=0.75 means that in 75% of cases, controls have a higher value, of e.g. grey matter proportion, than AD subjects (Cliff, 1993).

8.2.2.4.3 Comparing parametric and nonparametric effect sizes

The parametric and nonparametric effect sizes that I have described are measuring slightly different things: the parametric effect size is measuring the shift between distributions and the nonparametric effect size is measuring overlap between distributions. However, these can both be transformed to the probability that a randomly selected subject from the control group will have a higher value than a randomly selected AD subject; referred to as the "common language effect size statistic (CLES)" (McGraw and Wong, 1992; Coe, 2002).

In fact, there is no transformation required for the nonparametric effect size, it is directly calculated and is just the left side of equation 17, i.e. $\#(x_i>x_j)/nm$, where all values of x in group i are compared with all values of x in group j, and mn is the total number of comparisons (number of subjects in group i × number of subjects in group j).

The parametrically derived CLES is the cumulative Gaussian distribution function of *Cohen's d*, i.e. the *P*–value in the Gaussian distribution table (Freedman et al., 2007) that corresponds to the calculated *Cohen's d* (*Cohen's d* is equivalent to a *Z*–score of the Gaussian distribution). Providing the data are Gaussian distributed, the parametric CLES should be approximately equal to the nonparametric CLES (McGraw and Wong, 1992; Vargha and Delaney, 2000; Coe, 2002).

8.2.2.5 Lobular reporting of voxel-wise statistics

Because of the apparently lobular weighted pattern of atrophy in ageing and AD, e.g. there are thought to be relatively weak effects in the parietal lobe, partial effects in the frontal lobe, and strong effects in the temporal lobe (e.g. Thompson et al., 2003), I report all voxel-wise statistics by lobe. That is, I segmented the aged standard space into the four major lobes: frontal, temporal, parietal, and occipital (Figure 42). I did this using atlas-based segmentation (chapter 5) with the "Hammers-mith" atlas (Hammers et al., 2003).

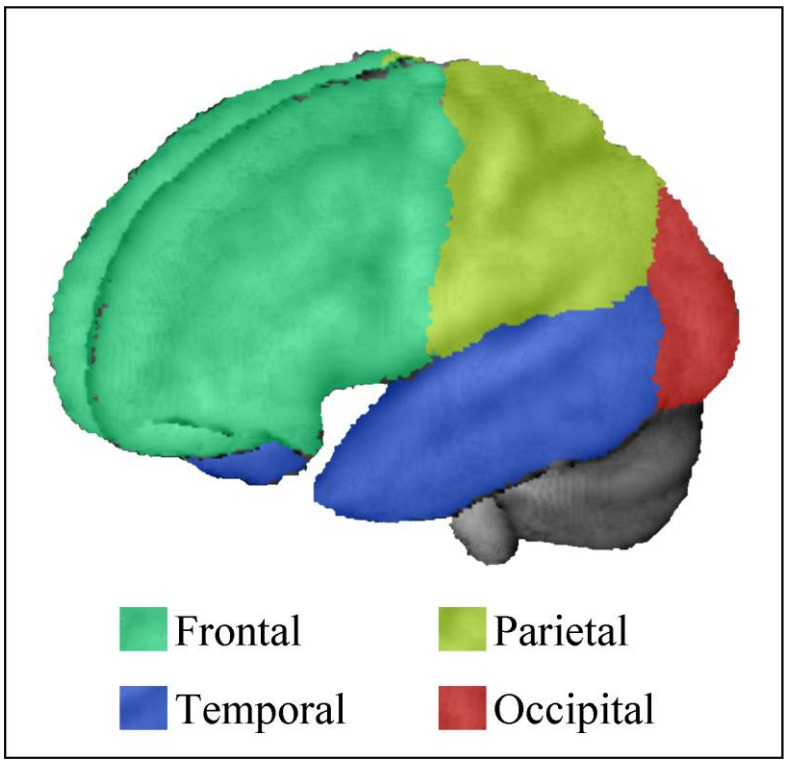


Figure 42. Segmentation of aged standard space into the four major lobes: frontal, temporal, parietal, and occipital
These lobes were segmented via atlas-based segmentation using the "Hammers-mith" atlas (Hammers et al., 2003) for the purpose of reporting voxel-wise statistics.

8.3 Results

Before constructing parametric and nonparametric reference atlases, I assessed the voxel–wise distributions of grey matter proportion in the reference atlas subjects.

8.3.1 Assessments of voxel–wise grey matter distributions in the normal reference atlas subjects

Examples of grey matter proportion distributions in the cortex and subcortical regions are shown in Figure 43.

To assess grey matter distributions throughout the cortex I calculated the kurtosis and skewness at each voxel (Figure 44).

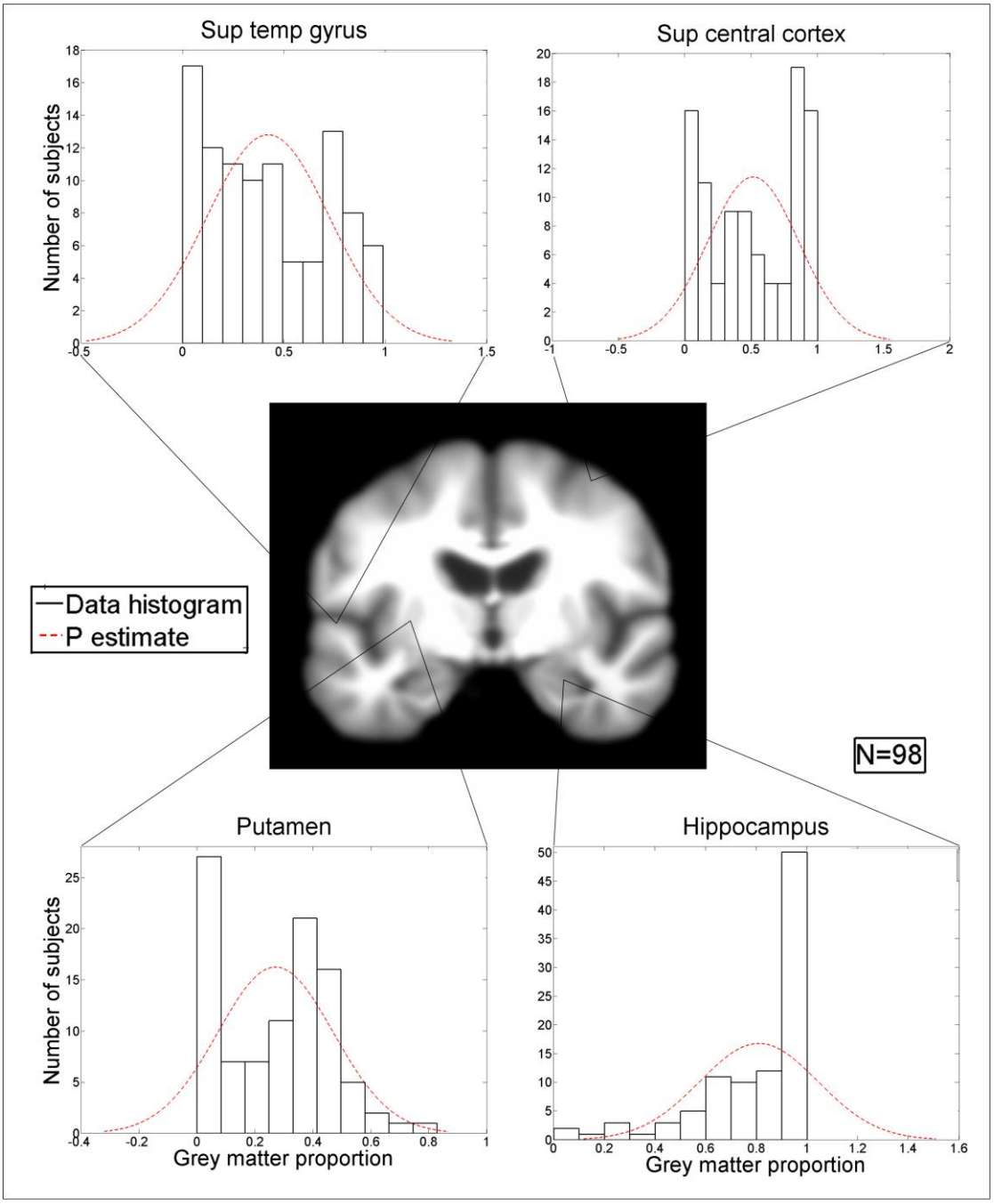


Figure 43. Grey matter proportion distributions in 98 normal aged subjects (60–90 years) at single voxels in the superior temporal gyrus (top left), superior central cortex (top right), putamen (bottom left), and hippocampus (bottom right)
These distributions are markedly different to the assumed parametric (P) distribution (red dashed line).

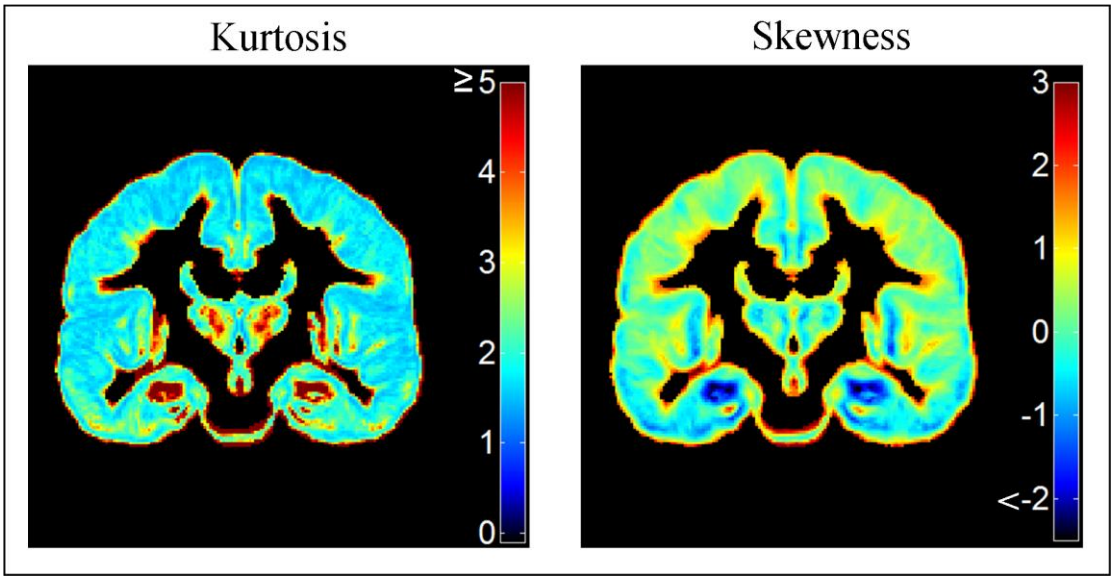


Figure 44. Voxel–wise kurtosis and skewness in the grey matter proportion atlas subjects
The Gaussian distribution has kurtosis of 3 (yellow) and skewness of 0 (light green/blue).

Figure 44 shows that kurtosis and negative skewness was greatest in the hippocampal region and Table 30 shows that grey matter proportion was not Gaussian distributed in much of the cortex (Gaussian distributed data have kurtosis of 3 and skewness of 0).

Table 30. Median kurtosis and skewness in voxels in the grey matter proportion atlas subjects

	Kurtosis	Skewness
(*median, IQR*)	1.95, 0.95	0.11, 0.92

As much of the cortex was not Gaussian distributed (Figure 44 and Table 30), I constructed parametric and nonparametric atlases of grey matter proportion.

8.3.2 Parametric and nonparametric atlases of grey matter proportion

The parametric and nonparametric atlases of grey matter proportion are shown in Figure 45.

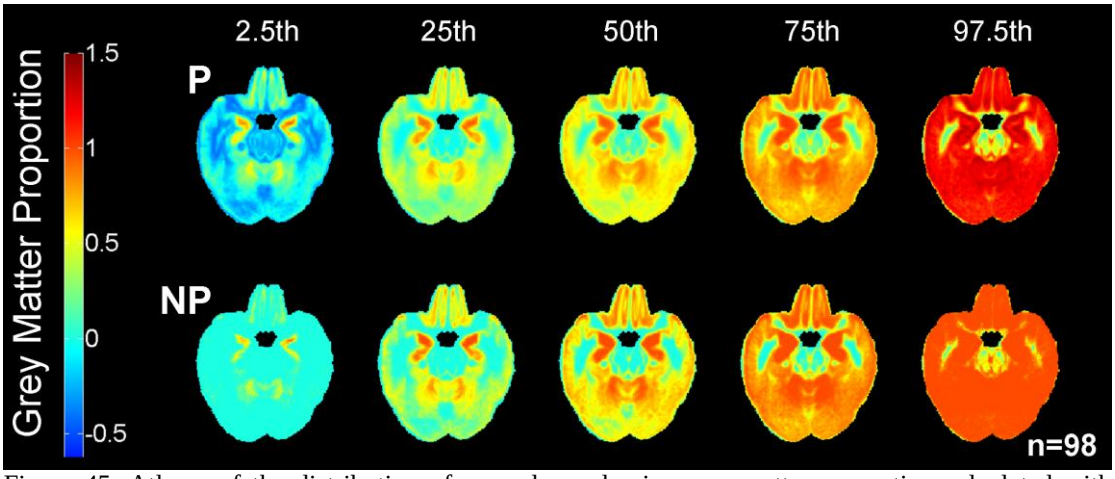


Figure 45. Atlases of the distribution of normal voxel–wise grey matter proportion calculated with parametric (P – upper panel) and nonparametric (NP – lower panel) statistics in 98 aged normal subjects (60–90 years)
The parametric atlas frequently overestimates the lower normal limits of grey matter (dark blue negative proportions) and the upper limits (dark red proportions greater than 1). The ranking method used in nonparametric statistics means that it does not make these errors. This z–slice location shows the hippocampal region.

As shown in Figure 45, the parametric method determined that the lower clinical limit of grey matter proportion in many voxels was less than zero and that the upper clinical limit was greater than one. This does not make sense mathematically or physiologically. The lower clinical limit of any voxel in the nonparametric atlas was not less than zero and the upper clinical limit was never greater than one. Further differences between the atlases are illustrated by the histograms of voxel values in Figure 46.

The visually apparent differences between the parametric and nonparametric atlases (Figure 45 and Figure 46) are quantified in Table 31. Although they are often thought to be approximately equal (Freedman et al., 2007), the mean (parametric 50^{th} percentile) and median (nonparametric 50^{th} percentile) histograms had similar kurtosis but differed in skewness by more than 200%, i.e. the parametric method generally overestimated the central tendency of grey matter proportion. The methods were in most agreement at the 75^{th} percentile but even here the parametric method generally overestimated grey matter proportion by 52%. As a reminder, in commonly used brain imaging analyses, e.g. the GLM, parametrically derived estimates for percentile ranks are assumed to equal the nonparametrically derived values (Freedman et al., 2007).

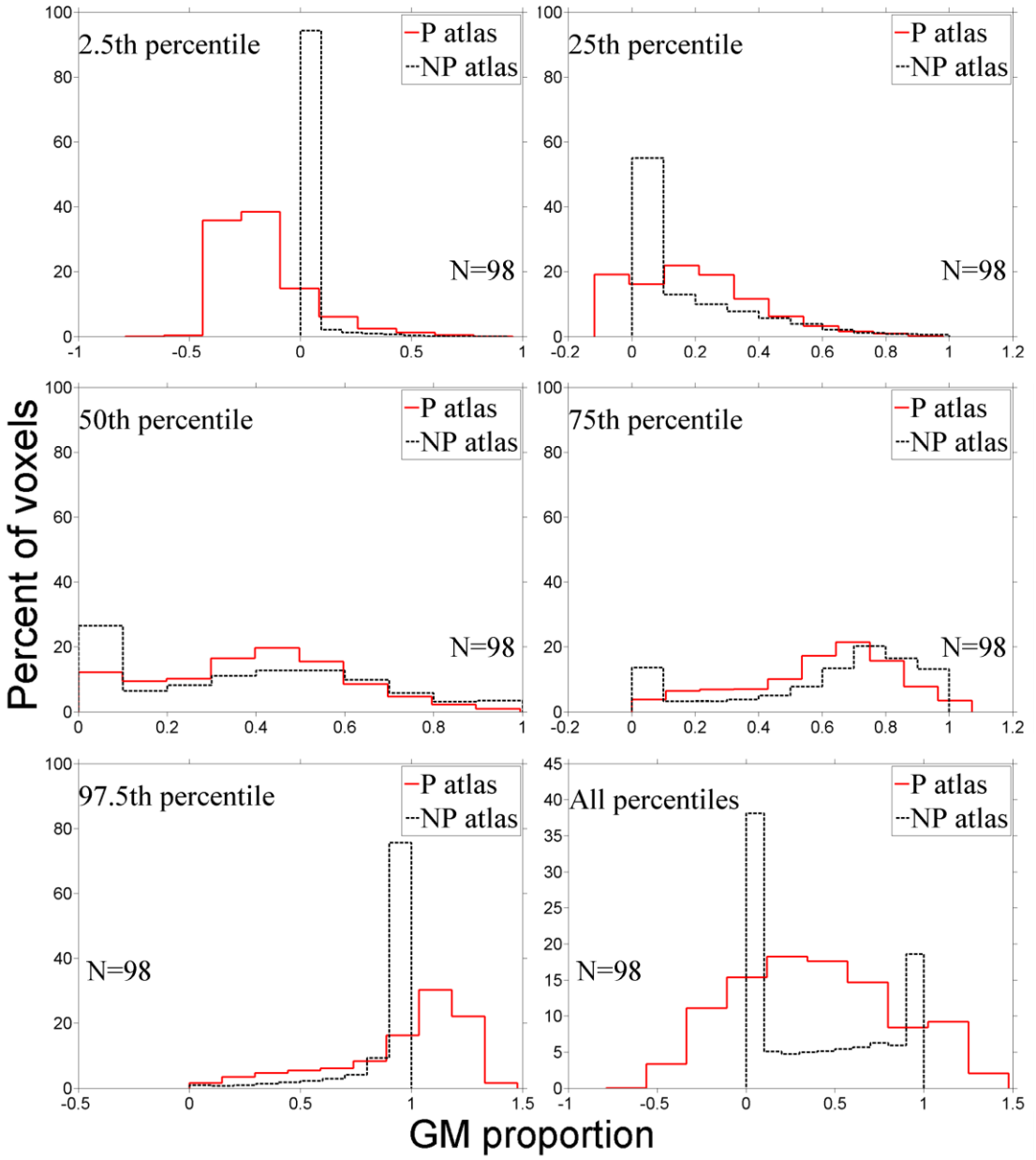


Figure 46. Histograms of the parametric and nonparametric grey matter atlas percentile levels

The greatest difference between the atlases was at the lower clinical limit (2.5th percentile) and this led to the parametric atlas producing false classifications in many voxels when used to assess AD subjects.

Table 31. Kurtosis and skewness in parametric and nonparametric grey matter proportion atlas histograms

	Kurtosis			Skewness		
Percentile	*Parametric*	*Nonparametric*	*%Err*	*Parametric*	*Nonparametric*	*%Err*
2.5th	5.99	33.76	–464	1.55	5.27	–240
25th	3.29	4.63	–41	0.75	1.45	–93
50th	2.45	2.04	17	0.06	0.20	–233
75th	2.53	2.42	4	–0.52	–0.79	–52
97.5th	3.43	10.31	–201	–1.13	–2.69	–138

Note: %Err=percent error between parametric and nonparametric templates

8.3.3 The effect of parametric assumptions in voxel classification: false classifications

False classifications occur in parametric methods when data do not follow the Gaussian distribution: when the parametrically defined lower clinical limit (2.5th percentile) is less than the nonparametric limit it produces false negatives (false normals) and when the parametric limit is greater than the nonparametric limit it produces false positives (false abnormals) (Elveback et al., 1970; Freedman et al., 2007).

Table 32. Median percentage of voxel–wise false classifications in ADNI and OASIS with the parametric GM atlas

Classification	ADNI (*median, IQR*)	OASIS (*median, IQR*)
False negatives	32.2, 8.2%	34.1, 9.3%
False positives	48.4, 6.2%	45.5, 6.6%

Note: ADNI=Alzheimer's Disease NeuroImaging Initiative; OASIS=Open Access Series of Imaging Studies; IQR=interquartile range.

Table 32 was derived from the percentage of voxel–wise false classifications within each subject to determine the median percentage of voxel–wise false negatives and false positives that occurred in the ADNI and OASIS samples (with the parametric GM atlas). This shows that, for most subjects in both samples, the parametric atlas incorrectly classified ~25–45% of true positive (abnormal) voxels as normal (false negatives); ~40–50% of the voxels classified as abnormal by the parametric atlas were within the normal range (false positives).

In both samples there were concentrated regions of false negatives in the mid sagittal caudal and anterior regions (Figure 47). Throughout whole lobular structures, the proportions of false negatives were approximately equal between the frontal and temporal lobes (Table 33, rounding errors mean that not all rows in this table sum to exactly 1).

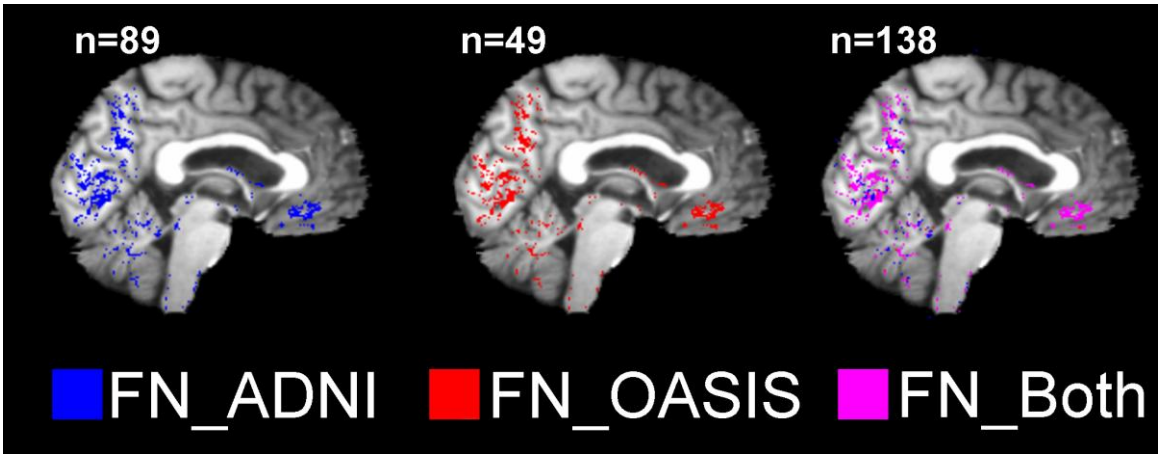


Figure 47. The locations of false negative/ false normal (FN) voxels in Alzheimer's disease (AD) subjects (aged 60–90 years) from ANDI (n=89) and OASIS (n=49) via voxel classification with the parametric grey matter atlas
The parametrically defined lower clinical limit (2.5th percentile) was less than the nonparametric limit (when it should have been the same) in these voxels and so defined these voxels as normal when in fact they were out with the normal range, i.e. abnormal.

Approximately two thirds of false positives occurred in the temporal lobe (Table 33), including the hippocampal region (Figure 48).

The concentrations of false classifications (Table 33, Figure 47 and Figure 48) suggest that parametric statistics may overestimate some differences, e.g. in the hippocampus, and underestimate other differences, e.g. in the frontal lobe, between normal ageing and Alzheimer's disease. Given the potential for misclassifications using the parametric method, I implemented voxel-based classification (and created the percentile rank atlas) using the nonparametric method.

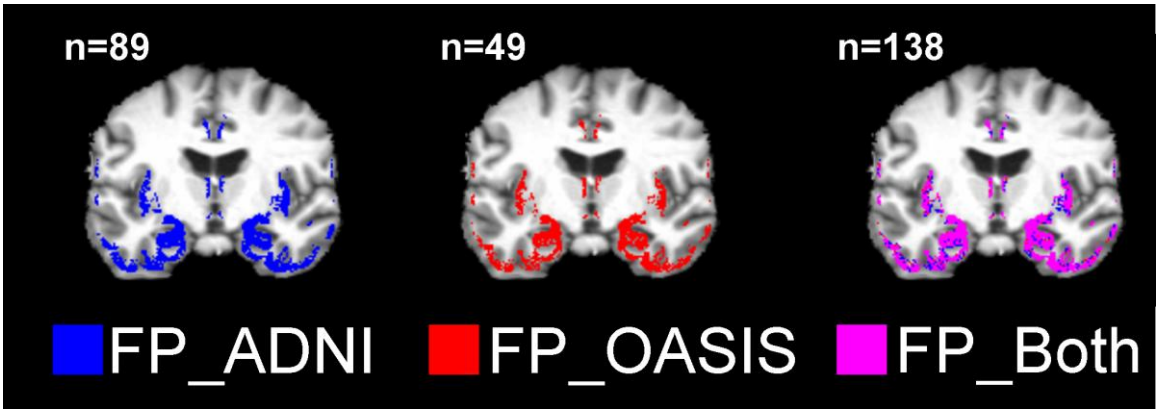


Figure 48. The locations of false positive/ false abnormal (FP) voxels in Alzheimer's disease (AD) subjects (aged 60–90 years) from ANDI (n=89) and OASIS (n=49) via voxel classification with the parametric grey matter atlas

The parametrically defined lower clinical limit (2.5th percentile) was higher than the nonparametric limit (when it should have been the same) in these voxels and so defined these voxels as abnormal when in fact they were still in the normal range.

Table 33. Proportions of misclassifications by lobe in ADNI and OASIS

		Frontal	Temporal	Parietal	Occipital
FN	ADNI	0.36	0.38	0.09	0.17
	OASIS	0.39	0.41	0.08	0.12
FP	ADNI	0.28	0.57	0.06	0.08
	OASIS	0.27	0.61	0.06	0.06

Note: FN=false negative; FP=false positive; ADNI=Alzheimer's Disease Neuroimaging Initiative; OASIS=Open Access Series of Imaging Studies.

8.3.4 A novel brain data transformation method: Percentile rank atlas-based ("clinical rank") transformation

Since the parametric template did not provide sufficiently close approximations of its assumed values, e.g. the minus two SD was often not equal to the 2.5th percentile, I used the nonparametric method to create the percentile rank atlas. Further to providing voxel-based classifications in individuals, this atlas may be used to meaningfully transform brain structure data for subsequent group-wise statistical analyses. So that they are representative of the group population, I first attempted to determine the number of subjects required to make a robust atlas.

8.3.4.1 Determining the number of subjects required for a representative percentile rank brain atlas

The representativeness of percentile rank atlases given their number of subjects are illustrated in Figure 49. This shows the percent similarity of percentile rank atlas histograms (with *n*=10, 20, …, 98 subjects) to the total *n*=98 percentile rank atlas histogram.

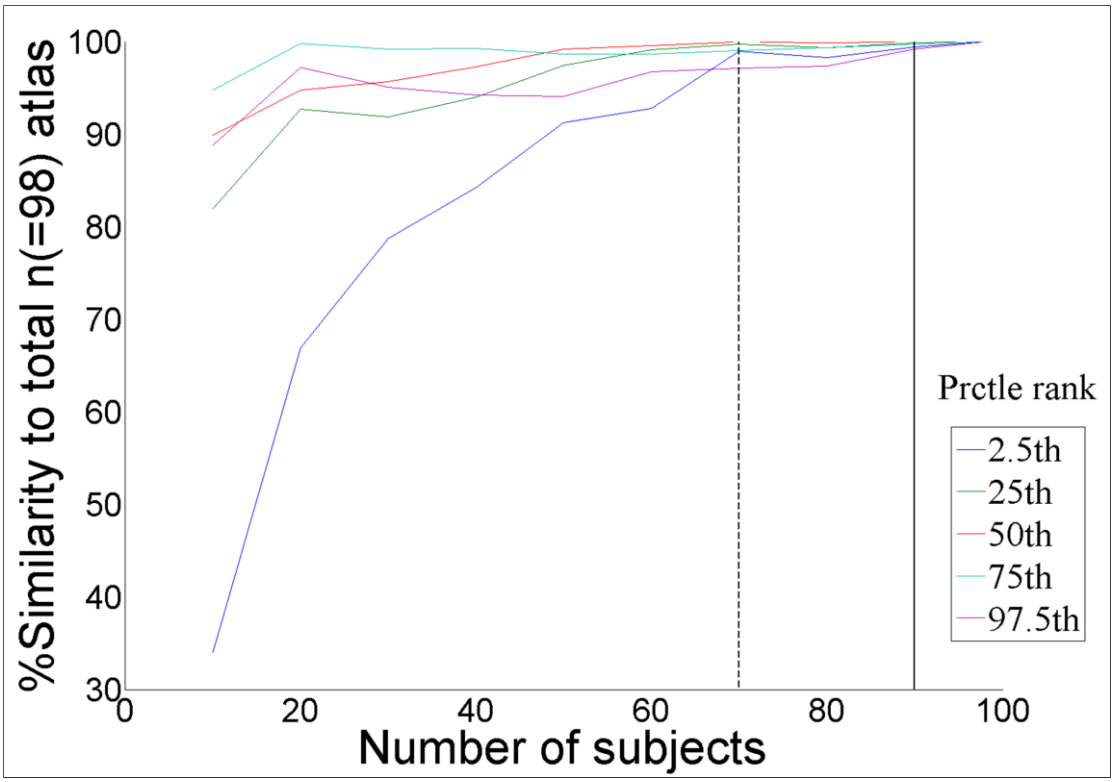


Figure 49. Determining the number of subjects required to recreate a complete percentile rank atlas representative of the total sample n=98 subjects
Each coloured line represents the change in each percentile (2.5^{th} – 97.5^{th}) given the addition of more subjects. Seventy subjects (~71%) were required to create an atlas that was 95% similar to the total *n*=98 atlas (referenced with the dashed vertical line) and 90 subjects (~92%) were required to create an atlas that was 99% similar to the total *n*=98 atlas (referenced with the solid vertical line).

For all percentile rank levels of the atlas, i.e. 2.5^{th}–97.5^{th} percentile ranks, 70 subjects (~71%) were required to create an atlas that was 95% similar to the total *n*=98 atlas. Ninety subjects (~92%) were required to create an atlas that was 99% similar to the total *n*=98 atlas. Each iteration of the atlas (e.g., 10, 20,… 70 subjects) was a random sample of the same 98 subjects. This information and repeated analysis will be useful when a much bigger sample of subjects is available, e.g. to determine a reasonable

amount of subjects to process when it is not feasible to process them all. For now I proceed with the total *n*=98 atlas and demonstrate how this may be used to meaningfully transform brain structure data so that it is suitable for subsequent statistical analysis.

8.3.4.2 Percentile rank atlas-based ("clinical rank") transformation

For many nonparametric comparisons between groups, e.g. Wilcoxon–Mann–Whitney, data distributions do not require to be Gaussian, however, they still require to be approximately equally shaped and have approximately equal variance (chapter 4). Figure 50 shows normal and AD subject distributions of grey matter proportion in a randomly selected voxel in the hippocampus.

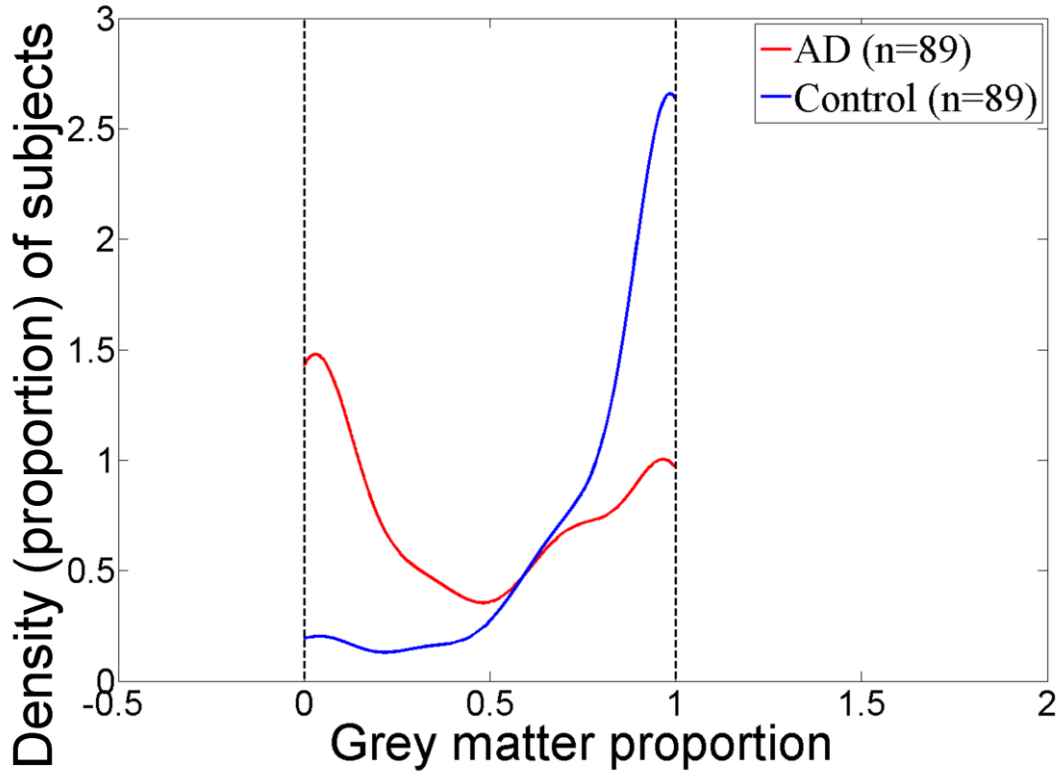


Figure 50. Distributions of grey matter proportion in a randomly selected voxel in the hippocampus in AD and control subjects
These raw data show that the distributions of these subjects are somewhat different.

Interquartile range (IQR, a measure of variance) of grey matter proportion was 0.2391 in the normal subjects and 0.8322 in the AD subjects. This meant that neither the distribution shapes nor variances between these groups were equal.

There are various data transformation methods available, e.g. logarithmic transformation, but these can be difficult to interpret and cause data to lose their meaning. By ranking grey matter proportion values in each subject in each group (according to the previously defined nonparametric reference for grey matter proportion; section 8.2.2.3), I transform these data into similarly shaped distributions but maintain a meaningful reference to the original data. Figure 51 shows an example of this "atlas-based clinical rank transformation" in GM from an individual AD subject. In this individual, the red region highlights the subtle hippocampal atrophy pathology associated with AD. This is consistent with their clinical diagnosis and suggests that NP voxel-based assessment may, in future, be a useful diagnostic tool.

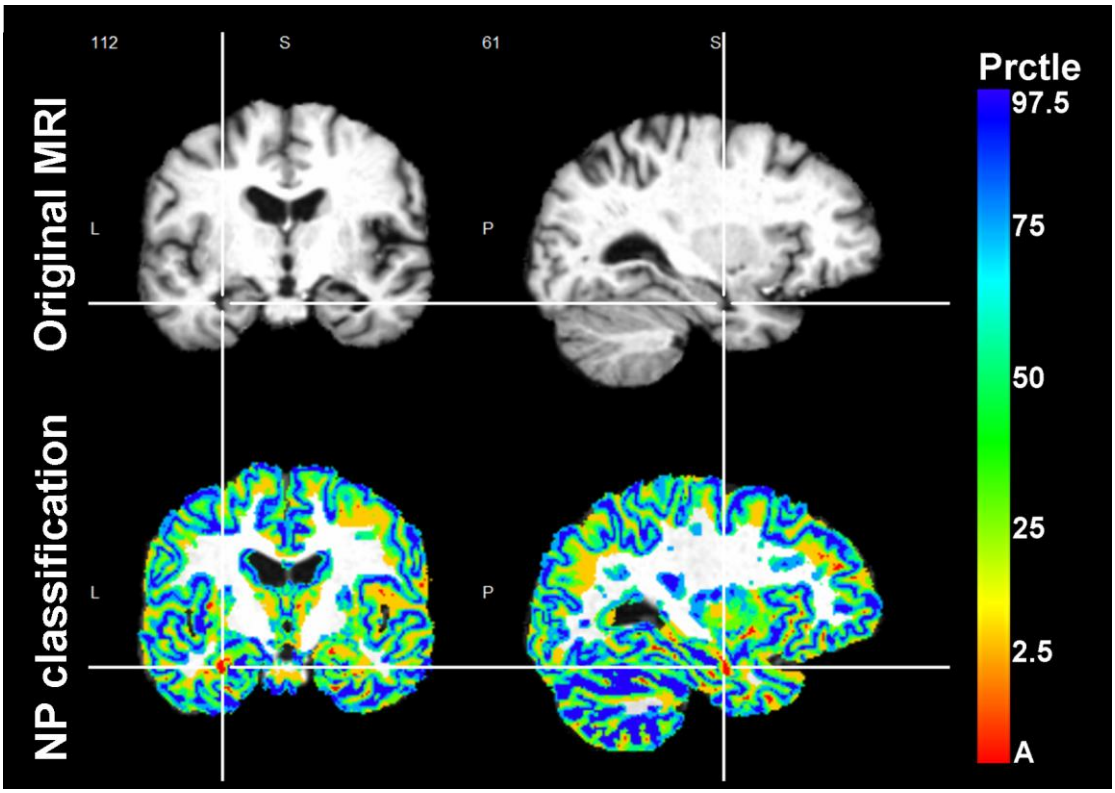


Figure 51. Nonparametric (NP) atlas-based ("clinical rank") transformation in GM from an individual AD subject.
Prctle=percentile; A=abnormal.

Each of the 89 AD and 89 normal subjects were ranked as in Figure 51. Figure 52 shows the distributions of ranks of grey matter proportion in these subjects (distributions of voxel values following atlas-based "clinical rank" transformation).

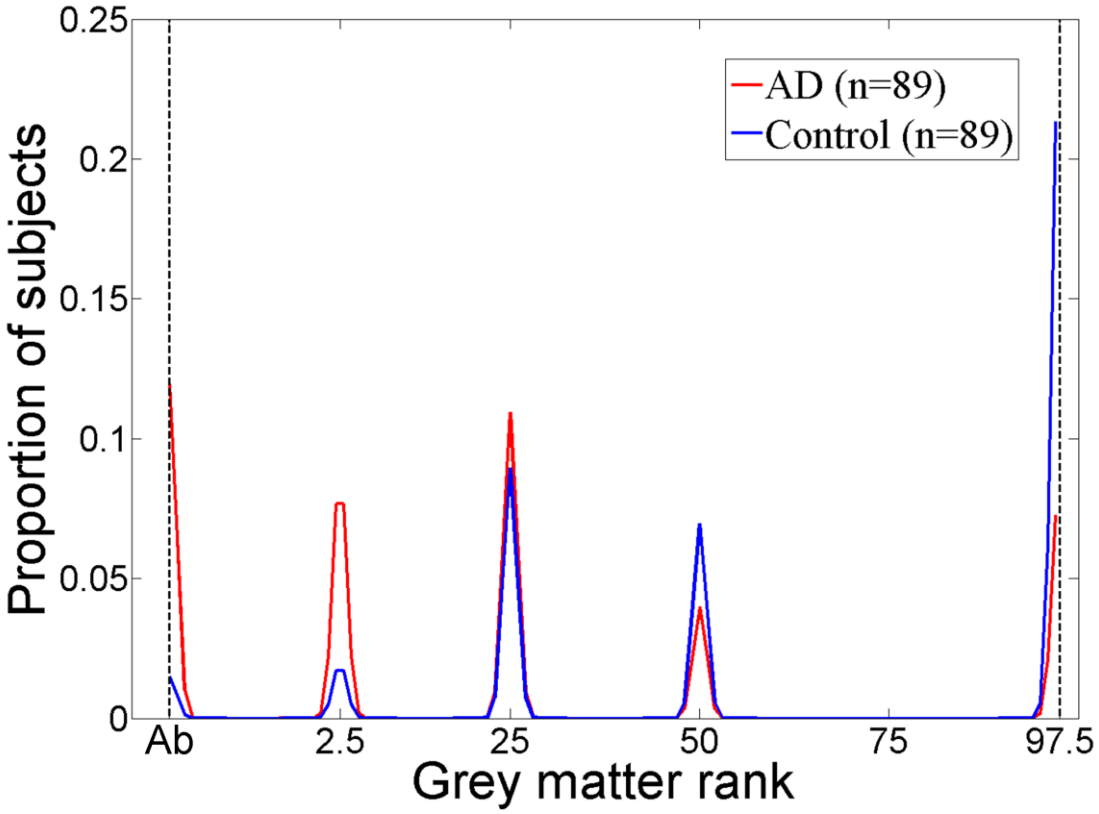


Figure 52. Distributions of grey matter ranks in the same randomly selected voxel in the hippocampus in AD and control subjects
This shows that the distributions of data in these groups now have a similar shape, a requirement for subsequent nonparametric, e.g. Wilcoxon-Mann-Whitney, testing. Ab=abnormal, i.e. less than the 2.5th percentile rank.

The histograms of both groups were in the same shape (frame of reference) following ranking but the original differences are generally maintained, e.g. the differences between groups at the limits of grey matter proportions and ranks. Further, the IQR of ranks of grey matter proportion was also similar after ranking: 72.5 in the normal subjects and 70 in the AD subjects.

This suggests that voxel–based ranking may support more reliable nonparametric testing between groups. But, as in any transformation of data, this should be carefully considered before use (Scott and Wild, 1991). In particular, grey matter proportion may take an infinite range of values between 0 and 1. I defined only

6 percentiles (including the abnormal, i.e. <2.5th, percentile) to rank these values. Therefore, while making the data more amenable to nonparametric statistical testing, clinically ranked image data may inherently have less information than the original image data.

A type of nonparametric effect size, "*delta*" (δ), does not require data to be normal, share distributions shapes, nor have equal variance (Cliff, 1993; Coe, 2002), i.e. does not require data transformation. Before implementing delta I tested the voxel–wise distributions of GM proportion to determine whether the more commonly applied parametric effect size was appropriate.

8.3.5 Effect sizes in group studies

To determine whether the parametric effect size was appropriate I first computed voxel–wise kurtosis and skewness. Kurtosis in Gaussian distributed data is equal to three and skewness is equal to zero. There was high kurtosis (≥5) and negative skewness (<–2) in the hippocampal region in both subject cohorts (ADNI Figure 53, OASIS Figure 54). Kurtosis and skewness by lobe in ADNI are shown in Table 34 and Table 35 (rounding errors mean that not all rows sum to exactly 1). Across both controls and AD subjects in ADNI, the temporal lobe consistently had the largest proportions of high kurtosis (>3.5) and negative skewness (<-0.5). The frontal lobe had the second largest proportions of these highly non Gaussian values.

Overall, from Table 34 and Table 35, only approximately 30% of the cortex in ADNI subjects was distributed approximately Gaussian.

Table 34. Proportions of voxel kurtosis by lobe in ADNI

	Kurt	k<2.5	2.5≤ k ≤3.5	k >3.5
	Lobe			
Controls	Frontal	0.73	0.26	0.02
	Temporal	0.59	0.35	0.06
	Parietal	0.76	0.24	0.01
	Occipital	0.75	0.25	0.00
AD	Frontal	0.72	0.27	0.01
	Temporal	0.67	0.30	0.04
	Parietal	0.71	0.28	0.01
	Occipital	0.73	0.27	0.00

Note: Kurt(k)=kurtosis;. The Gaussian distribution has kurtosis of 3.

Table 35. Proportions of voxel skewness by lobe in ADNI

	Skew	*s*<-0.5	-0.5≤ *s*≤0.5	*s* >0.5
	Lobe			
Controls	Frontal	0.09	0.67	0.25
	Temporal	0.29	0.56	0.16
	Parietal	0.03	0.65	0.32
	Occipital	0.04	0.68	0.28
AD	Frontal	0.07	0.66	0.28
	Temporal	0.19	0.63	0.18
	Parietal	0.02	0.59	0.40
	Occipital	0.02	0.68	0.31

Note: Skew(*s*)=skewness;. The Gaussian distribution has skewness of 0.

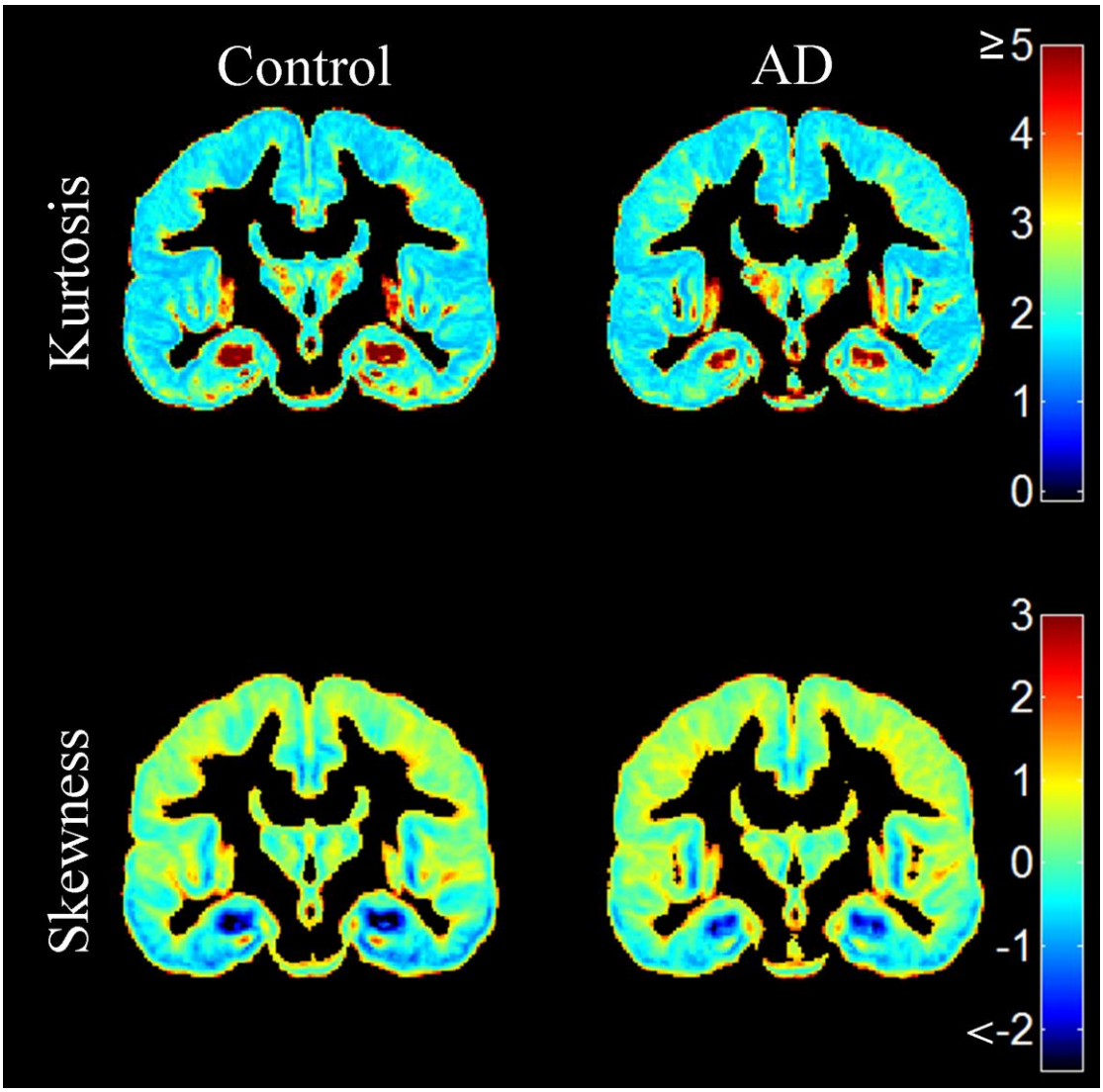


Figure 53. Voxel–wise kurtosis and skewness in the ADNI control (*n*=89) and AD (*n*=89) subject groups
The Gaussian distribution has kurtosis of 3 (yellow) and skewness of 0 (light green/blue).

I found a similar pattern of kurtosis and skewness in the OASIS data (Figure 54). Kurtosis and skewness by lobe in OASIS are shown in Table 36 and Table 37. As in ADNI, the temporal lobe consistently had the largest proportions of high kurtosis (>3.5) and negative skewness (<-0.5). The frontal lobe had the second largest proportions of these highly non Gaussian values. Again as in ADNI, only approximately 30% of the cortex in OASIS subjects was distributed approximately Gaussian.

Table 36. Proportions of voxel kurtosis by lobe in OASIS

	Kurt	$k<2.5$	$2.5\leq k\leq 3.5$	$k>3.5$
	Lobe			
Controls	Frontal	0.70	0.29	0.02
	Temporal	0.56	0.39	0.04
	Parietal	0.78	0.21	0.01
	Occipital	0.73	0.27	0.01
AD	Frontal	0.70	0.28	0.02
	Temporal	0.60	0.36	0.04
	Parietal	0.77	0.22	0.01
	Occipital	0.73	0.27	0.01

Note: Kurt(k)=kurtosis;. The Gaussian distribution has kurtosis of 3.

Table 37. Proportions of voxel skewness by lobe in OASIS

	Skew	$s<-0.5$	$-0.5\leq s\leq 0.5$	$s>0.5$
	Lobe			
Controls	Frontal	0.13	0.67	0.20
	Temporal	0.23	0.62	0.15
	Parietal	0.06	0.71	0.23
	Occipital	0.09	0.71	0.19
AD	Frontal	0.12	0.66	0.22
	Temporal	0.22	0.59	0.19
	Parietal	0.05	0.68	0.27
	Occipital	0.07	0.71	0.22

Note: Skew(s)=skewness;. The Gaussian distribution has skewness of 0.

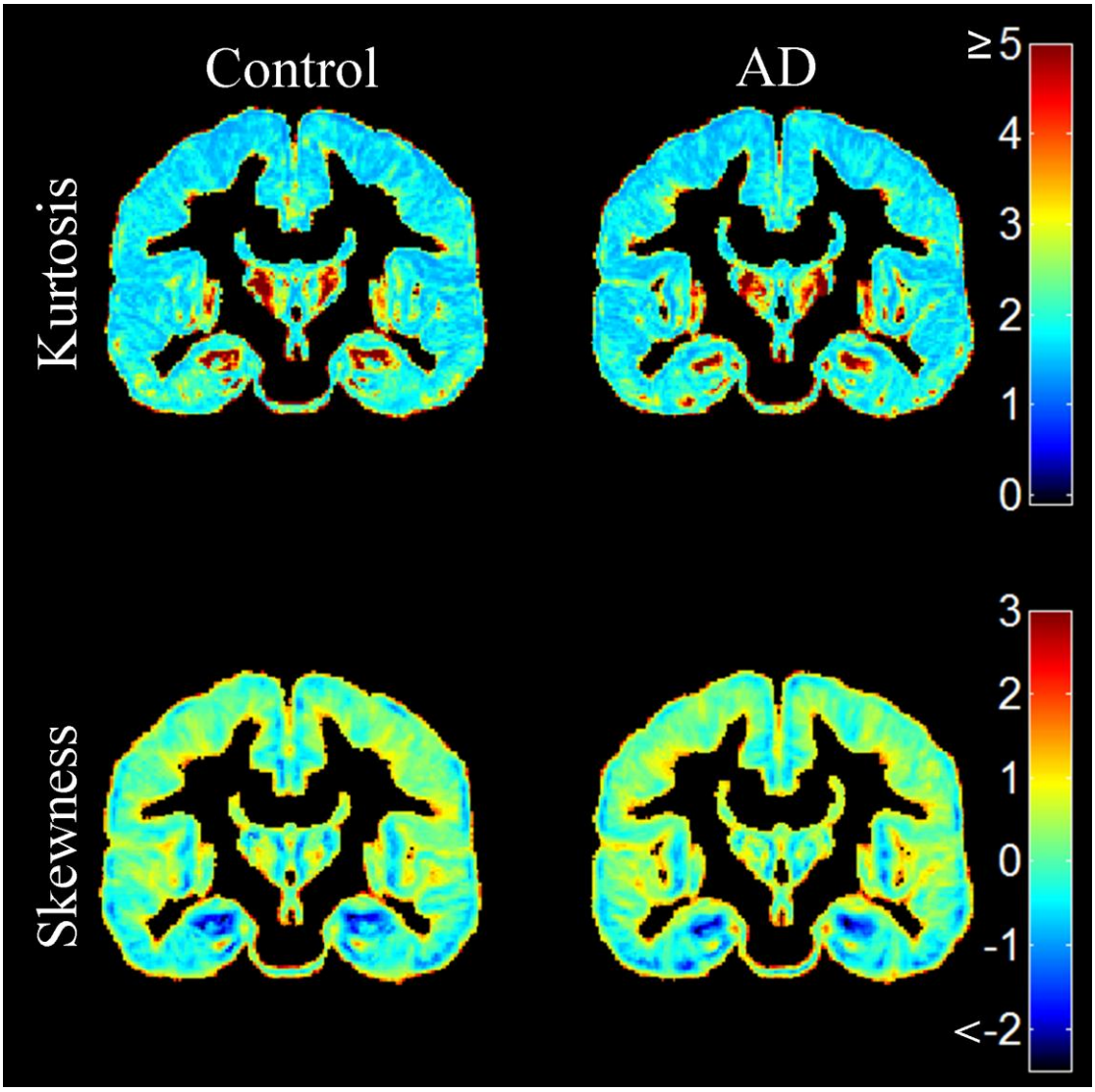


Figure 54. Voxel–wise kurtosis and skewness in the OASIS control (*n*=49) and AD (*n*=49) subject groups
The Gaussian distribution has kurtosis of 3 (yellow) and skewness of 0 (light green/blue).

Figure 53 and Figure 54 show that the largest deviations from the Gaussian distribution are in the frontal and temporal lobes, specifically in the hippocampal region.

This suggests that, especially in these areas, parametric effect size may not be a reliable indicator of differences between normal ageing and AD subjects. I assessed this by comparing parametric CLES with nonparametric CLES. If data are Gaussian, parametric CLES will be equal to nonparametric CLES (McGraw and Wong, 1992; Vargha and Delaney, 2000; Coe, 2002).

The CLES derived from parametric effect size (*Cohen's d*) and nonparametric effect size (*delta*, δ) between AD and control subjects in ADNI and OASIS are shown

in Figure 55 and Figure 56, respectively. It is apparent that the largest effects between groups are in the temporal lobe, particularly the hippocampus. However, the parametric CLES frequently overestimated the differences between normal ageing controls and AD in this region by up to 55% (Figure 57). As already discussed (in the methods section 8.2.2.4.3), if the data were Gaussian distributed, parametric CLES would have been approximately equal to nonparametric CLES (McGraw and Wong, 1992; Vargha and Delaney, 2000; Coe, 2002).

Parametric and nonparametric CLES in ADNI and OASIS are listed by lobe in Table 38 and Table 39 (rounding errors mean that not all rows sum to exactly 1). These show that as the number of subjects increases from 49 per group (OASIS) to 89 per group (ADNI), parametric and nonparametric methods tend to generally agree on the proportion of high effects (CLES≥71). This was apparent in all lobes with the exception of the temporal lobe. Differences between the actual sizes of effects determined by each method are listed by lobe in Table 40 (rounding errors mean that not all rows sum to exactly 1).

Across all lobes, the differences in CLES between methods were generally reduced with an increased number of subjects but even with 82% more subjects, agreement between the methods improved by no more than 10%. Although there was general agreement between methods in approximately 80% of the cortex, even with this increase in subjects, 20% of the cortex still exhibited marked differences (>10%) between methods. This means that parametric methods may produce misleading results, i.e. results that do not correspond to the meaning of effect size in figure 40, in a significant part (20%) of the cortex. The highest differences between methods (d ≥30%) within the frontal lobe were reduced, but not eliminated entirely with this increase in subjects. The highest differences remained within the temporal lobe regardless of the number of subjects. Although only ~2% of voxels exhibited these differences, Figure 55, Figure 56, and Figure 57 show that these were in the very significant hippocampal region.

Table 38. Proportions of Voxel CLES by lobe in ADNI

	CLES	$c<56$	$56\leq c<64$	$64\leq c<71$	$c\geq 71$
	Lobe				
Parametric	Frontal	0.61	0.35	0.04	0.00
	Temporal	0.46	0.39	0.11	0.04
	Parietal	0.63	0.33	0.03	0.00
	Occipital	0.62	0.34	0.03	0.00
Nonparametric	Frontal	0.90	0.10	0.00	0.00
	Temporal	0.75	0.22	0.02	0.01
	Parietal	0.91	0.09	0.00	0.00
	Occipital	0.88	0.12	0.00	0.00

Note: CLES(*c*)=common language effect size, the units of which are percent; These ranges of CLES were chosen as they correspond to the previously defined small (0.2/56), medium (0.5/64), and large (0.8/71) effect sizes.

Table 39. Proportions of Voxel CLES by lobe in OASIS

	CLES	$c<56$	$56\leq c<64$	$64\leq c<71$	$c\geq 71$
	Lobe				
Parametric	Frontal	0.53	0.38	0.08	0.01
	Temporal	0.49	0.39	0.10	0.02
	Parietal	0.54	0.38	0.07	0.01
	Occipital	0.53	0.38	0.08	0.01
Nonparametric	Frontal	0.87	0.12	0.01	0.00
	Temporal	0.83	0.15	0.02	0.00
	Parietal	0.85	0.14	0.01	0.00
	Occipital	0.84	0.15	0.01	0.00

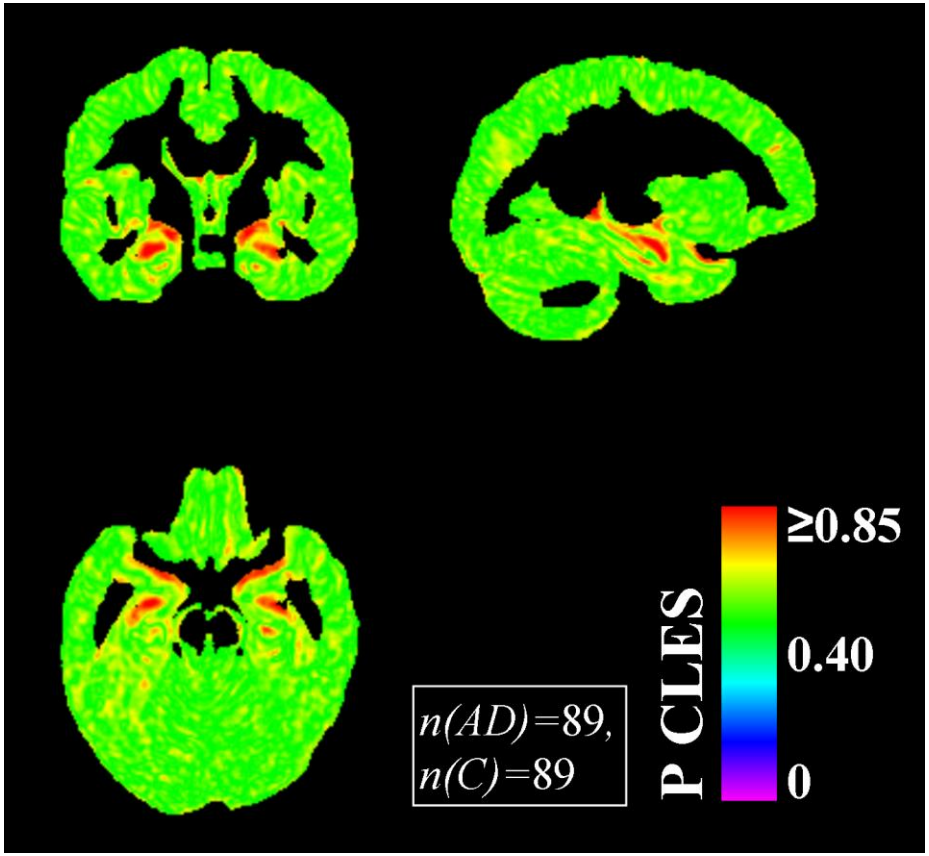
Note: CLES(*c*)=common language effect size, the units of which are percent; These ranges of CLES were chosen as they correspond to the previously defined small (0.2/56), medium (0.5/64), and large (0.8/71) effect sizes.

Table 40. Proportions of Voxel Parametric/ Nonparametric CLES differences by lobe in ADNI and OASIS

	Diff	$d<10$	$10\leq d<20$	$20\leq d<30$	$d\geq 30$
	Lobe				
ADNI	Frontal	0.79	0.17	0.04	0.01
	Temporal	0.83	0.12	0.03	0.02
	Parietal	0.83	0.15	0.02	0.00
	Occipital	0.82	0.15	0.03	0.00
OASIS	Frontal	0.72	0.19	0.07	0.02
	Temporal	0.73	0.19	0.07	0.02
	Parietal	0.76	0.18	0.05	0.01
	Occipital	0.76	0.17	0.05	0.01

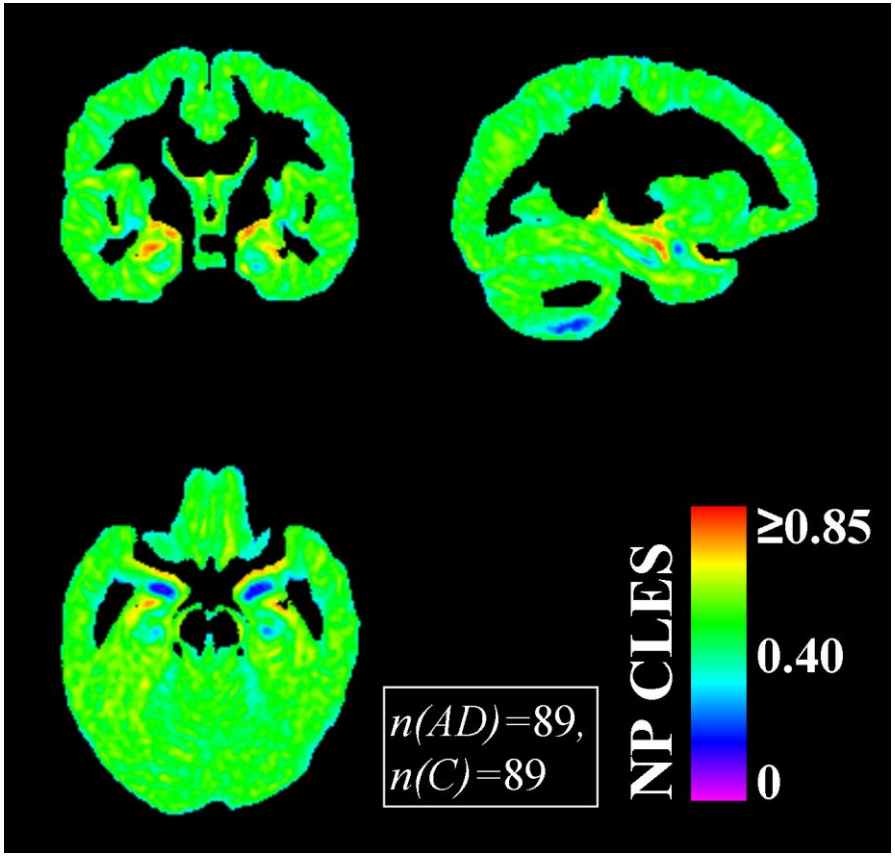
Note: Diff(d)=difference in common language effect size (measured in percent) between parametric and nonparametric methods.

PARAMETRIC CLES



i.

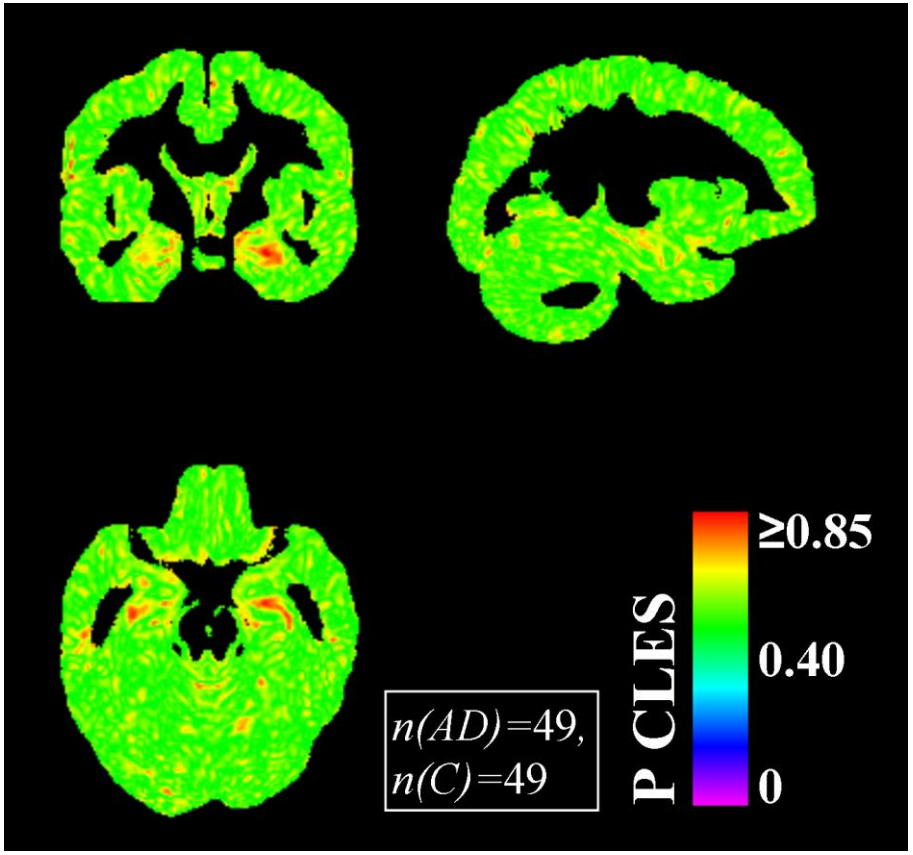
NONPARAMETRIC CLES



ii.

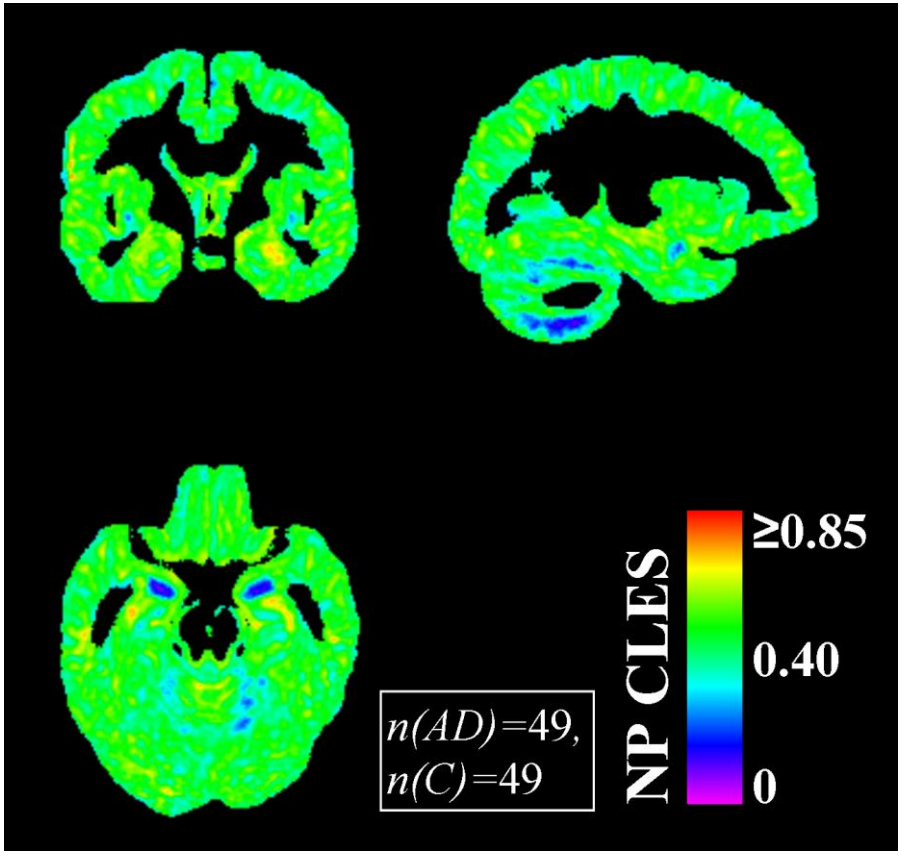
Figure 55. i. Parametric "common language effect size statistic" (P CLES) and ii. Nonparametric CLES (NP CLES) in the ADNI Alzheimer's disease (AD)–control (C) group study
ADNI=Alzheimer's Disease Neuroimaging Initiative.

PARAMETRIC CLES



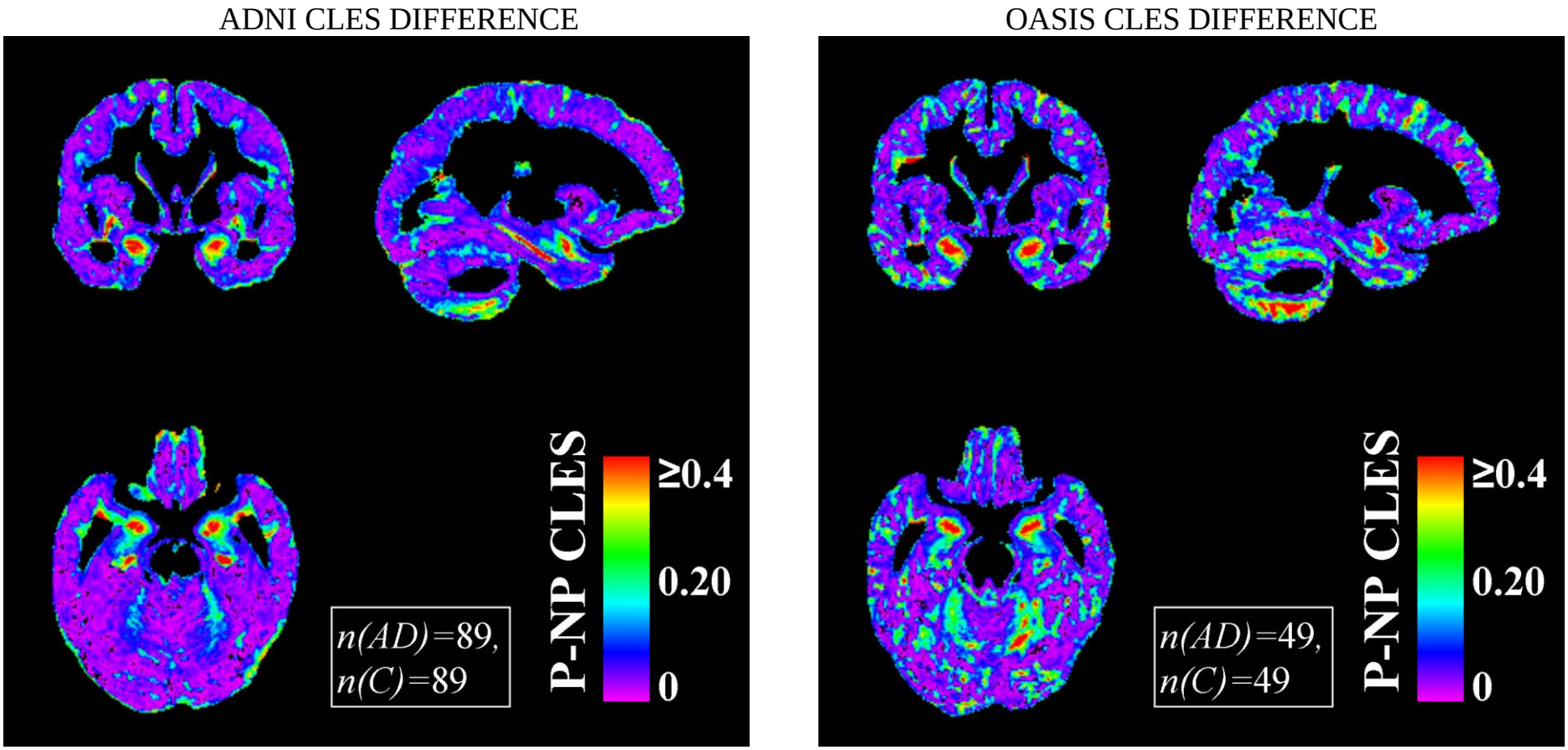
i.

NONPARAMETRIC CLES



ii.

Figure 56. i. Parametric "common language effect size statistic" (P CLES) and ii. Nonparametric CLES (NP CLES) in the OASIS Alzheimer's disease (AD)–control (C) group study
OASIS=Open Access Series of Imaging Studies.



i. ii.

Figure 57. Differences between parametric and nonparametric common language effect size (CLES) in the ADNI and OASIS Alzheimer's disease (AD)–control (C) group studies

Calculated by parametric (P) minus nonparametric (NP) CLES. ADNI=Alzheimer's Disease Neuroimaging Initiative; OASIS=Open Access Series of Imaging Studies.

Figure 55, Figure 56, and Figure 57 show that the separation between the distributions of GM in AD and controls may not be as extreme as suggested by the parametric effect size. The probability that a randomly selected subject from the control group will have a higher value than a randomly selected AD subject was artificially increased with the parametric method (by up to 40%; Figure 57).

I further tested the reliability of the parametric effect size by calculating standard differences at percentile ranks across the AD and control distributions (these should be approximately equal to the parametric effect size; Coe, 2002). The parametric and actual data distributions are shown in Figure 58.

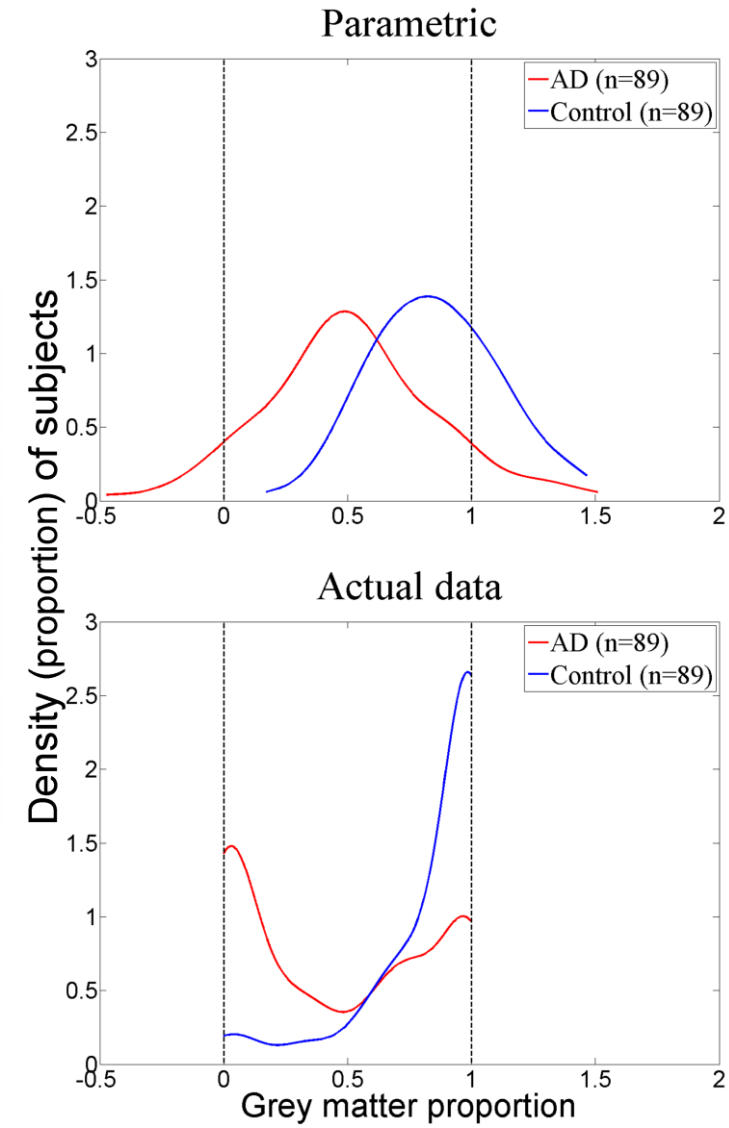


Figure 58. Example of grey matter proportion distributions in a randomly selected hippocampus voxel in AD and control subjects The parametrically assumed (upper panel) and actual distributions (lower panel) are shown.

Table 41. Parametric effect size and standard differences in a randomly selected hippocampus voxel between AD and control subjects

Parameter	AD	Control	Standard difference
Mean	0.4493	0.8444	1.1687*
2.5^{th}	0.0000	0.0021	0.0064
25^{th}	0.0247	0.7609	2.1777
50^{th}	0.3667	1.0000	1.8731
75^{th}	0.8569	1.0000	0.4232
97.5^{th}	1.0000	1.0000	0.0000

Note:*this is the overall parametric effect size ("*Cohen's d*").

The standard difference column in Table 41 shows that the parametric effect size cannot reliably be generalised across the actual data distributions, i.e. at some points in the distributions (2.5^{th}, 75^{th}, 97.5^{th} percentiles) the standard differences is much smaller than the effect size whereas it is much larger at others (25^{th} percentile). Although these data were from only one randomly selected voxel, Figure 53 and Figure 54 show that similarly shaped distributions were found throughout the hippocampus. I am not refuting that there are differences between AD and controls in the hippocampus – Figure 55, Figure 56, and Figure 58 show that there clearly are differences. It may be that these differences are not best described by parametric methods.

Further to the potential unreliability of parametric effect sizes, Figure 58 shows that parametric methods may inverse the actual data. The parametric distribution suggests that the majority of AD subjects have GM proportion about 0.5 whereas the inverse is true, the majority have either approximately 0 or 1 (this is an actual example of "the problem with Averages", that was vividly illustrated by Sir Francis Galton in 1889 – chapter 1). Furthermore, the parametric distribution leads to nonsensical grey matter proportion values, e.g. less than zero and greater than one. This provides more evidence that, when data are not Gaussian distributed, parametric statistics may not adequately describe the differences in brain structure between two groups, e.g. normal ageing subjects and subjects diagnosed with AD.

8.4 Discussion

The changes in brain structure that occur in neurodegenerative diseases, such as Alzheimer's, may be highlighted using a voxel-based normal reference (atlas). Many previous normal brain atlases were derived from mean (parametric) values (chapter 3). To reliably extrapolate these values to define the limits of normal brain structure, data must be distributed approximately Gaussian (Elveback et al., 1970; Freedman et al., 2007). I found that data were not distributed Gaussian in approximately 70% of the cortex (these data had either high kurtosis and/or negative skewness). Largest deviations from the Gaussian distribution were in the temporal and frontal lobes. The effect of this is that, if extrapolating mean values to define the limits of normal brain structure, parametric atlases produced many misclassifications (false positives and false negatives) in AD patients. A large percentage (60%) of false positives were in the temporal lobe, specifically focused in the hippocampus. This suggests that pathology in this region may be more subtle than expected by parametric atlases. Further, many (40%) of false negatives were in the frontal lobe which, although often not reported in research, is often implicated individual clinical diagnoses of AD (Geroldi et al., 1999; Thompson et al., 2003). My atlases therefore illustrate that the pathology of AD may not always be temporal lobe and hippocampus-centric, rather it may sometimes manifest in a fronto-temporal pattern.

The nonparametric atlas that I developed may be used when the Gaussian distribution assumption is not met. Further to classifying normal and abnormal voxel values, this atlas may be used to transform data in preparation for subsequent nonparametric analyses. By classifying data with this atlas data are transformed to have a similar distribution shape and IQR. This is a requirement of nonparametric analyses such as Wilcoxon-Mann-Whitney. Many data transformation methods currently exist, such as the commonly applied Box-Cox (Box and Cox, 1964; Kruggel, 2006). However, my ranking method may provide a meaningful reference to the original data, i.e. it describes the "clinical normality" of voxel values. While it provides this meaningful framework, my rank-based transformation method does not make data more Gaussian so should not be used for subsequent parametric analyses. Further, as with any transformation method, this should be carefully considered before use (Scott and Wild, 1991). In particular, grey matter proportion may take an infinite range of values between 0 and 1. I defined only 6 percentiles (including the abnormal,

i.e. <2.5th, percentile) to rank these values. Therefore, while making the data more amenable to nonparametric statistical testing, ranked image data may inherently have less information than the original image data. This loss of information is true for any image transformation, e.g. voxel-smoothing. But despite this, I have still provided a potentially useful and meaningful way of preparing data for nonparametric statistical analyses of normal and diseased groups.

Further to assessments of individuals, I tested parametric and nonparametric methods for assessing normal and diseased groups. Specifically, I implemented parametric and nonparametric measures of the CLES. Prior to this, I tested the distributions of two independent cohorts of normal and AD subjects. In both cohorts, I again found that much of the cortex was not Gaussian distributed (high kurtosis and negative skewness were again found in the temporal lobe, specifically the hippocampus, and frontal lobe). The result of this was that, even with an 82% increase in the number of subjects (from 49 to 89), parametric CLES provided markedly different results in approximately 20% of the cortex. For parametric measures of effect size to be meaningful, they should be approximately equal to the nonparametric measures (McGraw and Wong, 1992; Cliff, 1993; Vargha and Delaney, 2000; Coe, 2002). Following the increase in number of subjects, the highest differences between methods (≥30%) were reduced but not eliminated within the frontal lobe; while these differences remained within the temporal lobe regardless of the number of subjects. Although only ~2% of voxels exhibited these differences, these were in the very significant hippocampal region. Taken in consideration with the misclassifications I found in individuals, this suggests that nonparametric (rather than parametric) methods should be implemented when creating statistical models to support diagnoses of AD.

By using the mean ±2 (rather than exactly 1.96) SD, I may have slightly overestimated parametric limits but the associated error in classifications was 0.03%. I defined normal brain structure limits for subjects aged between 60 and 90 years. There are large differences in brain structure between these ages (Ge et al., 2002; Resnick et al., 2003; Farrell et al., 2009; Ferguson et al., 2010). Future work will create separate atlases for each year of life. Further, I will incorporate other clinical and cognitive data into these templates, e.g. atlases for specific blood pressures and/ or cognitive ability.

Given the reasonably well defined lobular pattern of atrophy in ageing and cognitive decline (Braak and Braak, 1997; Thompson et al., 2003; Fox and Schott, 2004), I reported voxel-wise statistics by atlas-based segmentation of the major lobes (frontal, temporal, parietal, and occipital). This segmentation was not manually edited and therefore not perfect. Further, I could not tabulate voxel-wise statistics specifically for subregional volumes within these lobes, e.g. the hippocampus. While I automatically segmented these volumes in younger subjects, this process was not reliable in aged subjects. Despite this, I still provided qualitative assessments of the hippocampus and future work will use semi-automatic segmentation to support reporting of statistics at the subregional level.

Further to automated subregional volume segmentations, I did not implement other commonly applied image processing methods. In particular, I did not use conventional nonlinear registration (I used Nsurf, which was described in chapter 5), nor did I apply voxel smoothing. Smoothing is used to reduce noise in images and also to make the distributions of intensities more closely follow the Gaussian distribution (Ashburner and Friston, 2000; Good et al., 2001). By essentially averaging the intensities of adjoining voxels, smoothing may mask subtle differences between health and disease (Shenton et al., 2001). Moreover this is a subjective process and there are no quantitatively derived optimal parameters (Zhao et al., 2012).

This means, however, that I do not know how the distributions of voxels will behave after conventional nonlinear registration and voxel smoothing. Indeed, my finding that brain structure is least Gaussian in the hippocampus is in conflict with a previous study that showed approximate concordance with their hippocampus data and the Gaussian distribution (Bonilha et al., 2006; Rorden et al., 2007). The 85 subjects in that study were much younger (mean age ~31 years, SD ~8.0; range: 17–60), and they used a different processing method (SPM) that implemented voxel smoothing. One reason for voxel smoothing is to make data more Gaussian (Ashburner and Friston, 2000; Good et al., 2001). I have presented a method that does not require data to be Gaussian. This may be a useful alternative for others when they cannot justify the use of smoothing. For example, differences between normal and neurologically diseased groups are often subtle and these may be removed by smoothing (Shenton et al., 2001; Fotenos et al., 2005).

I illustrated my individual and group assessments using grey matter proportion images from normal older subjects and subjects diagnosed with AD. However, the

software I developed may be applied to any image sequence from any group, e.g. it can produce a T2–weighted percentile rank atlas from normal young adults for assessing brain structure in individuals and groups with schizophrenia or a T1–weighted percentile rank atlas from full terms infants to assess brain structure in preterm individuals and groups.

Finally, I attempted to validate the results from this chapter by repeating my analyses in separate, independent samples. While I found relative agreement, these data are not sufficient to make reliable population inferences. Despite 70 subjects providing a reasonable approximation of the 5 percentile ranks I defined (2.5^{th}, 25^{th}, 50^{th}, 75^{th}, 97.5^{th}), this will likely not be sufficient to define the "true" values of all 100 whole number percentile ranks. At least 100 subjects, i.e. 1 subject per percentile, would be required for this. Even then, repeated nonparametric versus parametric analyses in multiple samples would be required to determine "true" population values.

However, at the time of my analyses, these were the only publicly available brain image data from subjects that were cognitively and medically representative of normal ageing. Given this lack of data, we are developing the Brain Images of Normal Subjects (BRAINS) databank, which is described in the next chapter.

9. The Brain Images of Normal Subjects (BRAINS) Bank and Atlases

9.1 Introduction

The analyses in previous chapters (and those of others, e.g. Manolio et al., 1994; Ge et al., 2002; Resnick et al., 2003; Kruggel, 2006; Farrell et al., 2009) have shown that large amounts of data will be required to better understand the range of brain structure changes that occur through the lifecourse and neurological disease. For example, I found considerable overlap in brain volumes between normal ageing and AD subjects (chapter 6). Multivariate metrics, e.g. WMH volumes, from a range of MRI sequences will therefore be required to better differentiate these subjects. Brain MRI databanks have the facility to store the required data and the derived atlases of brain structure that may highlight pathology associated with neurological disorder (chapter 8).

As described in chapter 2, banks of brain data have been in development since at least the 1980s (Foulkes et al., 1988) and these banks started to include images from the 1990s (Cocosco et al., 1997). This meant that several publicly accessible brain MRI databanks existed prior to the present work, notably the ADNI, Human Connectome Project, IXI, ICBM, MIRIAD, and OASIS databanks. Although they serve their own purposes very well, these banks may be limited by a lack of: (1) medical metadata, e.g. blood pressure and cognitive tests, to support the classification of "normal" subjects; (2) subjects that represent the characteristics of cognitively normal older people (>60 years), e.g. high blood pressure; (3) demographic diversity, e.g. most subjects were acquired in American centres; (4) MR image sequences, e.g. FLAIR and T2–weighted images for segmentation of potential lesions and venous sinuses; and (5) robust, i.e. nonparametric, atlases of brain structure.

In this chapter I describe the "Brain Images of Normal Subjects (BRAINS)" bank and the atlases I intend to derive from these data. My atlases will be used to investigate the range of brain structure in normal ageing and to highlight pathology of neurological diseases. After describing the BRAINS bank infrastructure, I describe development of: (1) "normal mean difference" assessment in my preterm infant subjects using the publicly available MNI normal infant atlas; (2) an "adultspan" (25–90 year) atlas of brain structure that I created from ADNI, OASIS, and our early to middle adult subjects; and (3) an "aged" (60–90 year) atlas that I created from ADNI and OASIS data and used to highlight pathology associated with AD.

9.2 BRAINS bank methods

9.2.1 Subjects

The BRAINS bank will initially include 1595 different normal subjects aged <1–80 years (~50% female). These were acquired in 10 separate studies, specifically to assess normal brain development and ageing or as controls for studies of neurological disease (Table 42).

Table 42. BRAINS bank MRI data

Study Name	N	Age[1]	Tesla	Sequences
Preterm infant controls	10	<1	3	T1, T2, DTI
Bipolar controls (McIntosh et al., 2008)	102[L]	16–25	1.5	T1, T2
High risk controls (Job et al., 2002)	179[L]	16–25	1, 1.5	T1, T2
Psychiatry controls (McIntosh et al., 2007)	70[L]	20–50	1.5	T1, T2
CaliBrain (Moorhead et al., 2009)	15[M]	25–42	1.5	T1
Amygdala controls (Hall et al., 2008)	100	25–45	1.5	T1, T2
NIH DTI	80	25–64	1.5	T1, T2, T2*, MT–MRI, FLAIR, DTI
Normal Ageing Brain (MacLullich et al., 2002)	150[L]	65–75	1.5, 2	T1, T2, T2*, FLAIR
LBC 1936 (Deary et al., 2007)	729[L]	70–75	1.5	T1, T2, T2*, FLAIR, DTI
LBC 1921 (Shenkin et al., 2003)	160 [L]	75–90	2, 1.5	T1, T2, T2*, FLAIR, DTI
All	1595	<1–80		

Note: [1]Age is in years; L=longitudinal data; M=multicentre data; NIH DTI=National Institutes of Health Diffusion Tensor Imaging; LBC=Lothian Birth Cohort.

Longitudinal data were acquired for at least 1119 subjects (data collection is ongoing) and multicentre data for 15 subjects (highlighted with an “L” and “M”, respectively in Table 42).

9.2.2 Medical assessment

The majority of subjects aged under 65 years were determined to be medically normal if they were not on any long term medication (with the exception of the contraceptive pill), not diagnosed with diabetes or high blood pressure, and had no history of diagnosed neurological disorder, major psychiatric disorder, substance misuse (including alcohol), or cranial surgery.

Medical history was also acquired in subjects over 65 years and further to this they received a physical examination (Deary et al., 2007). This examination recorded: height, weight, visual acuity in both eyes, time to walk six metres, ability to stand from sitting, demi–span, head circumference, activities of daily living (Townsend, 1979), blood pressure, peak expiratory flow rate in the lungs, forced expiratory volume in one second, forced vital capacity, grip strength in both hands, and date of the menopause in women. We have limited data on the ethnicity of subjects.

Following medical tests and examinations, each subject completed a battery of cognitive tests.

9.2.3 Cognitive assessment

To determine that they were cognitively normal, all subjects under 65 years (except the infants) completed at least a subset of: the Wechsler Adult Intelligence Scale III (WAIS-III; Wechsler, 1997a), Wechsler Memory Scale III (WMS-III; Wechsler, 1997b), National Adult Reading Test (NART; Nelson and Willison, 1991), verbal fluency (Lezak et al., 2004), and simple and four–choice reaction time (Deary et al., 2001) tasks. Subjects used as controls in psychiatric studies were also assessed using the Structured Clinical Interview for DSM–IV (SCID).

Subjects aged over 65 years completed a similar battery of tests but further to these they completed the Hospital Anxiety and Depression Scale (Zigmond and Snaith, 1983) and Mini–Mental State Examination (MMSE; Folstein et al., 1975). Moreover, over 700 of the aged subjects have cognitive data that were acquired at 11 years (Deary et al., 2007). For subjects aged over 65 years, these are data unique in the world to BRAINS. After completing the medical and cognitive assessments, subjects that were within normal limits received brain MRI.

9.2.4 Brain MRI

The majority of brain MRI data were acquired in the Brain Research Imaging Centre, The University of Edinburgh with a GE Signa Horizon HDxt 1.5 T clinical scanner (General Electric, Milwaukee, WI). For most subjects the imaging protocol consisted of: axial T2–, T2*–, and FLAIR–weighted sequences, a coronal T1–weighted volume sequence, an axial T1–weighted fast–spoiled gradient echo (FSPGR) sequence, a magnetization transfer (MT–MRI) pulse sequence, and a diffusion MRI protocol, all acquired in DICOM. All image sequences were determined to be within normal appearing limits by a consultant neuroradiologist. The DICOM metadata and aforementioned cognitive and medical data were automatically catalogued to populate the databank structure.

9.3 Databank structure

The internally housed databank was built by Dr. Dominic Job and Dr. David Rodriguez Gonzalez, The University of Edinburgh, using the MySQL database software. We intend to make the databank publicly available using the Extensible Neuroimaging Archive Toolkit (XNAT) platform. XNAT is a widely used and customisable platform that will provide external users a portal to the internal databank. This portal will allow various levels of access that will determine what data users can access, e.g. data owners will have full access whereas unaffiliated users will have restricted access based on data owner specifications.

At the time of writing, the internal databank was still under construction. However, its general structure may be described briefly as follows: each subject is assigned a row in each metadata table (containing various columns of variables) and each of these records (rows) are linked via their anonymised ID number. Subject images are stored in a computer server external to the MySQL tables. Corresponding subject images are similarly linked to the metadata tables via their anonymised ID number. The schema (variables) used to create these tables are described in the following sections.

9.3.1 Metadata schema

Adding all of the studies together, there are 220 demographic and clinical variables (Appendix AII.1). These include blood pressure, ethnicity, pulse, body mass index (BMI), medications, and a wide range of cognitive tests, e.g. MMSE, NART. These data were not collected for all subjects, e.g. infant subjects did not complete the MMSE, but each subject had at least the data listed in Table 43. Cognitively normal status was established from the infants' mother.

Table 43. Minimum metadataset in BRAINS bank

Variable	Type
ID	Ordinal
Age	Continuous
Gender	Binary
MRI report	Qualitative
Cognitively normal	Binary
Medically normal	Binary

Further to these demographic and clinical metadata, we also stored MRI metadata (e.g., sequence name, image orientation) for each subject via an automatic cataloguing tool I developed.

9.3.2 MRI data cataloguing tools

9.3.2.1 Automatic cataloguing of DICOM header information (MRI metadata)

As well as subject specific information, e.g. age, gender, databank queries may be interested in image specific information, e.g. in a volumetric study T1–weighted images are likely to be needed while MT–MRI may not. To allow these queries I had to extract this information from the DICOM headers and insert them into a databank table. The large number of images contained in BRAINS (Table 42) meant that this would have taken a long time to do by hand (I would have needed to open 4785 DICOM headers by hand and then manually copy and paste the required information into a databank table). Further, this amount of manual work provided a high probability of human error.

I therefore wrote a program in the MATLAB computing language to automatically read DICOM headers and extract the required information to a databank table. The metadata I extracted from the DICOM headers are listed in Table 44. Some of the data I extracted were for quality control, e.g. ensuring correct age with acquisition and birth date, and will not be included in the public BRAINS databank.

Table 44. Metadata automatically extracted from DICOM headers

DICOM header field	Definition
StudyID[P]	Identification of each scan session, usually in numerical order from 1, e.g. the ID of the scanners first session is 1.
SeriesNumber[P]	Number of each scan sequence, usually in numerical order from 1, e.g. the localiser sequence is 1.
PatientBirthDate[P]	Patient date of birth
AcquisitionDate[P]	Date of scan session
MagneticFieldStrength	Magnet field strength of scanner in tesla (T)
SeriesDescription	Description of each sequence, e.g. “ax FLAIR” (axial FLAIR)
ImageOrientationPatient	Acquisition plane, e.g. sagittal =[~0;~1;~0;~0;~0;~1]
RepetitionTime	Time between sequence pulses
EchoTime	Time between sequence pulse and maximum signal
Width	In–plane image width
Height	In–plane image height
SliceThickness	In–plane image thickness
SpacingBetweenSlices	Space between in–plane images

Note: P=private data, will not be included in the public BRAINS databank.

After extracting these metadata we converted the DICOM images into NIFTI–1 format using MRIcron (Rorden, 2010).

9.3.2.2 MRI format conversion

For computational analysis of brain MRI, DICOM format images are often converted to NIFTI–1 (as I did here). However, neuroradiologists may wish to view the results of computational analyses, e.g. voxel–based brain ranking (chapter 8), beside original images on their DICOM format workstations. To allow this I wrote a program in the MATLAB computing language to convert NIFTI–1 images back to DICOM. Since NIFTI–1 does not contain all of the DICOM header fields, I wrote a program to reinsert the DICOM header information lost in the initial NIFTI–1 conversion.

We chose to provide images in the NIFTI–1 format because it is the most commonly used image format among other established brain MRI databanks. I was able to ascertain this and other information about established brain MRI databanks through a series of visits I made (Table 45).

9.3.3 Links with established brain MRI databanks

To improve utility of the BRAINS bank, I sought to create collaborations with existing brain MRI databank centres. With a "Postdoctoral and Early Career Researcher Exchange" award from the Scottish Funding Council, I visited three of the worldwide leading brain MRI databank centres (Table 45). The aim of these visits was to initiate a worldwide brain MRI databank and atlas of brain structure across the lifecourse. By bringing together data from centres across the world, this databank and atlas may provide better understanding of normal brain structure variation and potentially diagnoses of neurological disorders.

9.3.3.1 Centres visited

The centres and principal investigators visited during the exchange are listed in Table 45.

Table 45. Centres and principal investigators visited

University	Centre	PI	Visit date
McGill University	Montreal Neurological Institute (MNI)	Alan C. Evans	30/03/13–06//04/13
University of California, Los Angeles (UCLA)	Laboratory of NeuroImaging (LONI)	Arthur Toga	06/04/13–12/04/13
Washington University in St. Louis	Neuroinformatics Research Group (NRG)	Daniel Marcus	14/04/13–21/04/13
Northwestern University	Neuroimaging and Applied Computational Anatomy Lab	Lei Wang	17/04/13

9.3.3.2 Results

As a result of these visits I acquired additional resources, data, and methods to develop the BRAINS bank and atlas. As described in the following list, I:

1. Was granted access to global computing and image processing resources (http://cbrain.mcgill.ca/)
2. Learned the MNI (http://www.bic.mni.mcgill.ca/ServicesSoftware/CIVET – Figure 59) and LONI (http://BrainSuite.loni.ucla.edu/ – Figure 60) image processing pipelines to
    a. Significantly reduce my processing times (by 50%, see section 5.4.1.3)
    b. Measure cortical thickness (previously I was only able to measure volume)
3. Learned methods to extract brain tissue volumes from infant subjects
4. Secured technical support for public release of the BRAINS bank on the widely used XNAT platform (https://central.xnat.org/)
5. Increased our sample size by ~1800 subjects

Preliminary results from the analyses conducted using the MNI and LONI pipelines are shown in Figure 59 and Figure 60, respectively. I used the MNI pipeline to calculate cortical thickness in the 80 normal younger to middle adult subjects (chapter 5). I then calculated the lower normal limit (2.5^{th} percentile) of cortical thickness using the parametric and nonparametric methods described in chapters 7 and 8. The blue areas in Figure 59 show where parametric were less than the nonparametric limits and the red areas where the parametric were greater than the nonparametric limits (the green areas show where the methods were equal).

For one of the young to middle adulthood subjects, I also computed cortical thickness using the LONI pipeline (Figure 60).

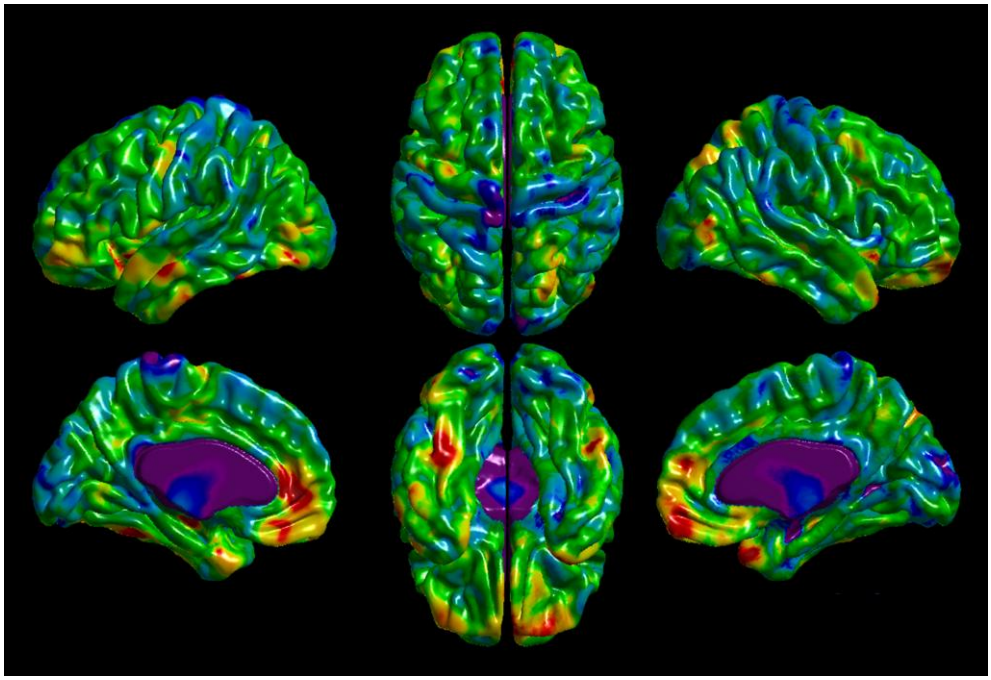


Figure 59. Illustration of the work I completed at MNI
Normal lower limit (2.5th percentile rank) of cortical thickness across adulthood (80 normal subjects aged 25–64 years) calculated with parametric and nonparametric methods. The red areas show where the parametric method underestimated the normal lower limit and the blue areas where the parametric method overestimated the normal lower limit; the green areas show where the limits from the parametric and nonparametric methods were approximately equal.

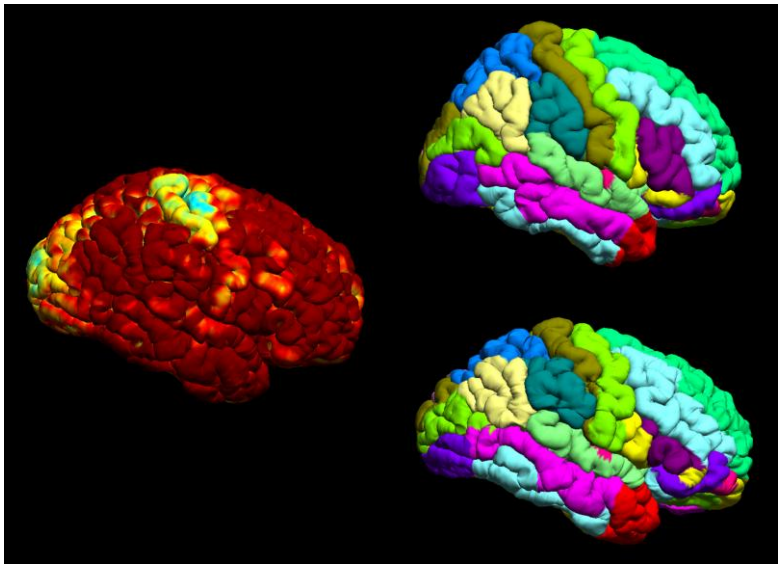


Figure 60. Illustration of the work I completed at LONI
Cortical thickness in an individual subject aged 25 years (left), measured using the LONI pipeline. The blue areas are ~2mm and the darkest red areas are ≥5mm. The LONI atlas (upper right) and resulting segmentation in the subject (lower right).

To measure reproducibility across methods, future work may compute cortical thickness in the same group of subjects and compare the results from MNI and LONI pipelines. As well as incorporating these methods and data from established banks, we will expand BRAINS in future by adding more local data.

9.3.4 BRAINS expansion

Further to the normal subjects that have already been recruited in The University of Edinburgh, we will integrate data from: all future normal subjects collected in Edinburgh; normal subjects already recruited in other centres across Scotland, e.g. University of Glasgow; and abnormal subjects already collected in Edinburgh, e.g. dementia patients, preterm infants. These data will allow us and others to test whether my atlases are useful for image processing, statistical analyses, and detection of pathology.

I am developing a series of BRAINS atlases and methods that may be useful for qualitative and quantitative image analyses. The following sections describe ongoing work into infant, “adultspan”, “aged” brain atlases.

9.4 Future directions

9.4.1 Infant BRAINS assessment

At the time of writing, infant control subjects were still being recruited (these have proved a challenge to recruit as compared to the preterm subjects). However, once a sufficient number have been collected I will develop a normative percentile rank atlas from these and use this in a similar way to the older adult percentile rank atlas, i.e. to highlight potential pathology in preterm subjects.

For present purposes I calculated percent differences between voxel values in the MNI normal infant atlas and voxel values in each preterm subject. This procedure is described in the following section.

9.4.1.1 Methods

To determine the normalcy of voxel values in preterm infant subjects (their demographics were described in section 5.2.1), I calculated percent differences from the term equivalent MNI normal infant atlas (Figure 61).

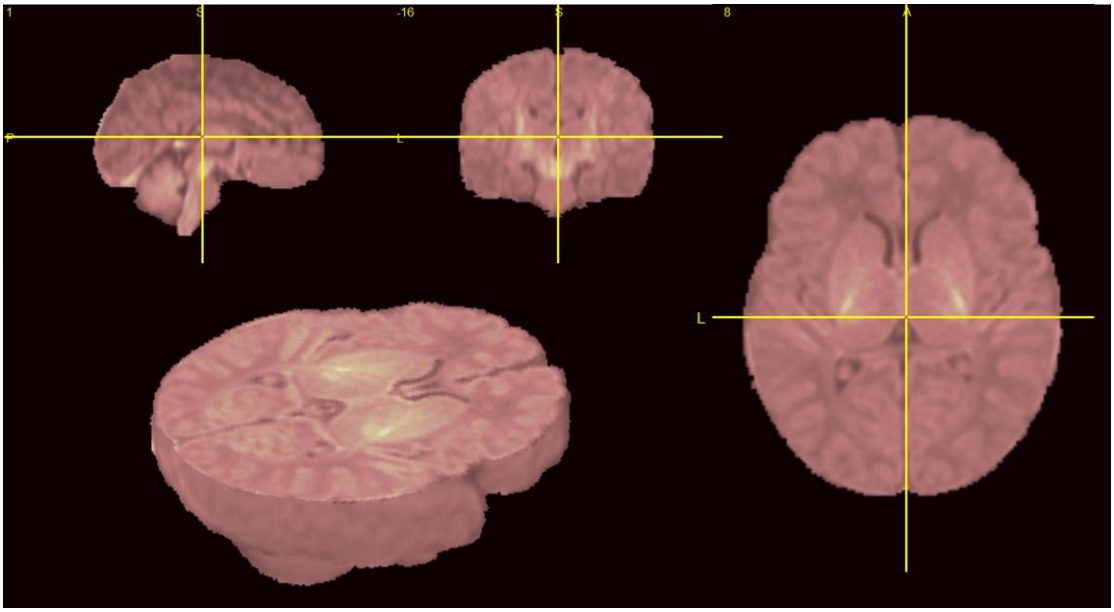


Figure 61. MNI normal infant atlas
This is the mean atlas derived from normal subjects at term equivalent age to the preterm subjects described in chapter 5. I used this to create "mean distance" images for the preterms.

Prior to calculating percent differences, I processed each subject according to the detailed procedures in chapter 5. Briefly, each subject was brain extracted and registered to the atlas space using Nsurf registration. The MR intensity range of each subject was then normalised to the atlas range using the SIS procedure, also described in detail in chapter 5. The intensity standardisation did not appear to unduly alter brain structure however I could only validate this qualitatively.

I did not perform N4 bias correction on the infant subjects prior to comparing them to the atlas because this appeared to unduly alter the developing deep grey matter structure, e.g., the contrast between the deep grey matter and surrounding white matter was reduced (Figure 62).

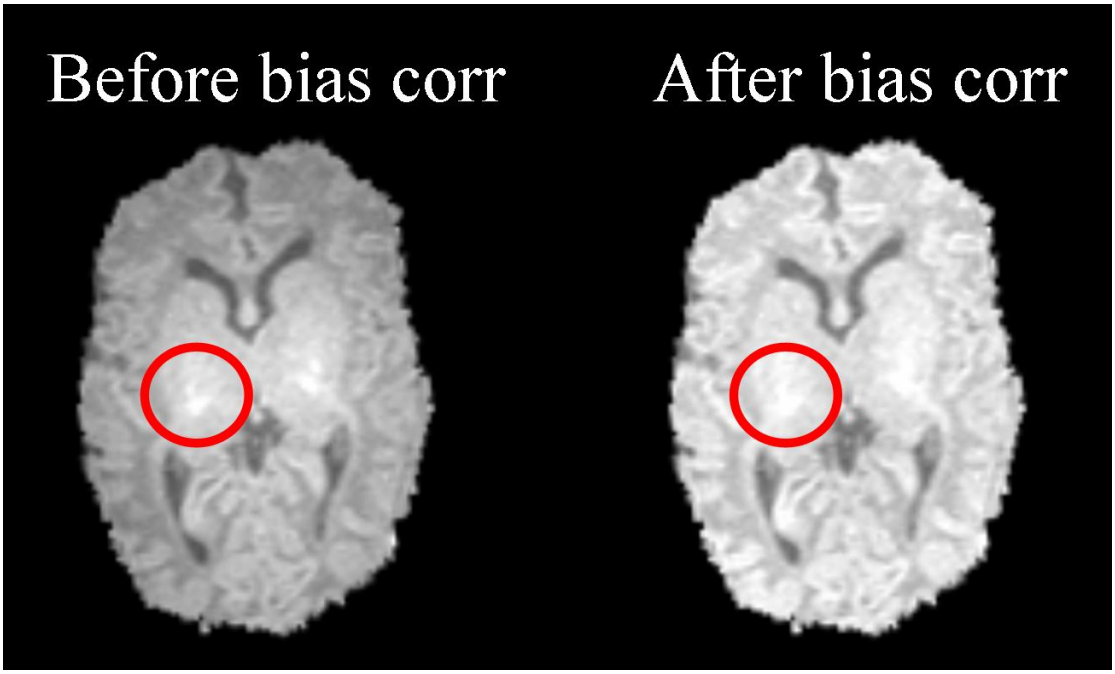


Figure 62. The effect of bias correction (corr) on a preterm infant subject
The contrast between deep grey matter and white matter was reduced after bias correction (highlighted by the red circles).

Following these processing steps percent differences between all voxel values in the atlas and all voxel values in each subject were calculated with equation 18,

$$\frac{\mu_i - x_{ij}}{\mu_i} \times 100 \qquad (18)$$

where μ_i is voxel i in the atlas, and x_{ij} is voxel i in subject j.

9.4.1.2 Results

The results from SIS in the preterm subjects are shown in Figure 63. This shows that the voxel intensity range of each of the 29 subjects was initially varied but normalised to the atlas range following SIS. Qualitative assessment of the images indicated that there were no undue alterations to contrast or brain structure in any subject following SIS.

The majority (75.9%, n=22/29) of preterm subjects had markedly greater than normal interhemispheric fissure cerebrospinal fluid volume, and more still (89.7%, n=26/29) had either increased interhemispheric fissure CSF or greater than normal inferior cerebellum CFS volume (red areas in Figure 64). Calculated quantitatively

with the normal reference atlas, these observations are consistent with radiological diagnoses of an experienced consultant neonatologist (Dr James Boardman, Royal Infirmary of Edinburgh). Moreover, they illustrate the high degree of skewness in preterm infant brain structure.

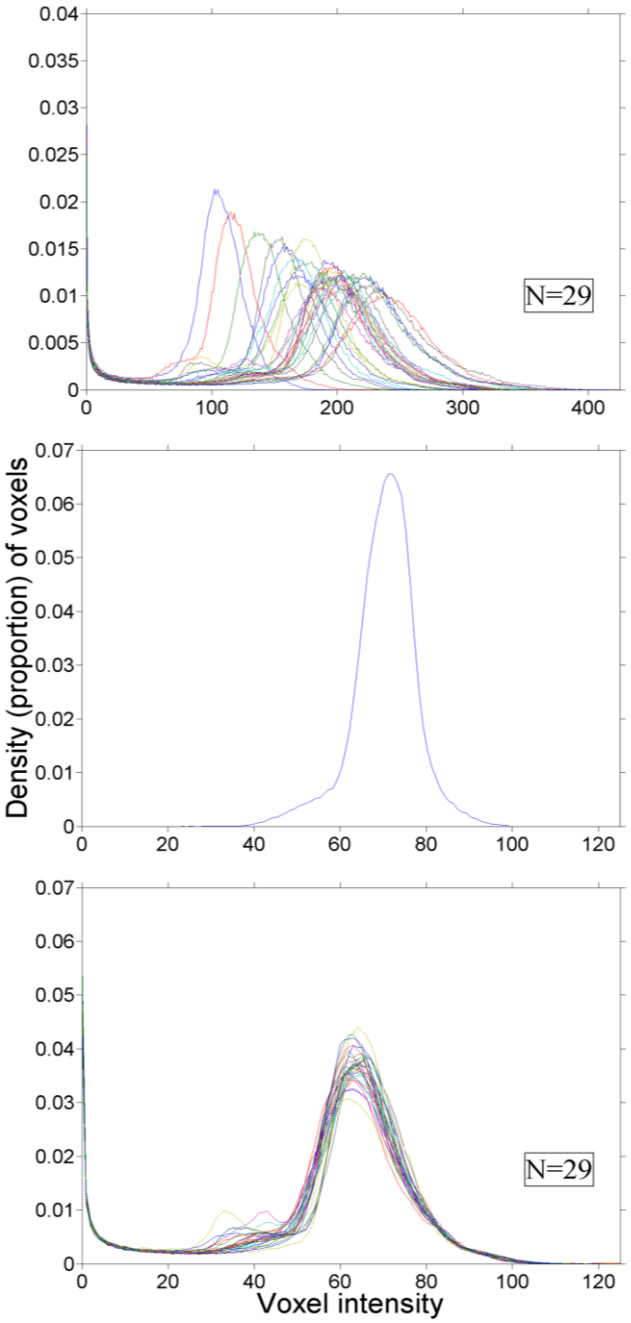


Figure 63. Standardised intensity scale (SIS) algorithm in the preterm subjects
Original subject intensity ranges (top panel) were transformed into the template range (middle panel) to produce standardised intensity ranges (bottom panel).

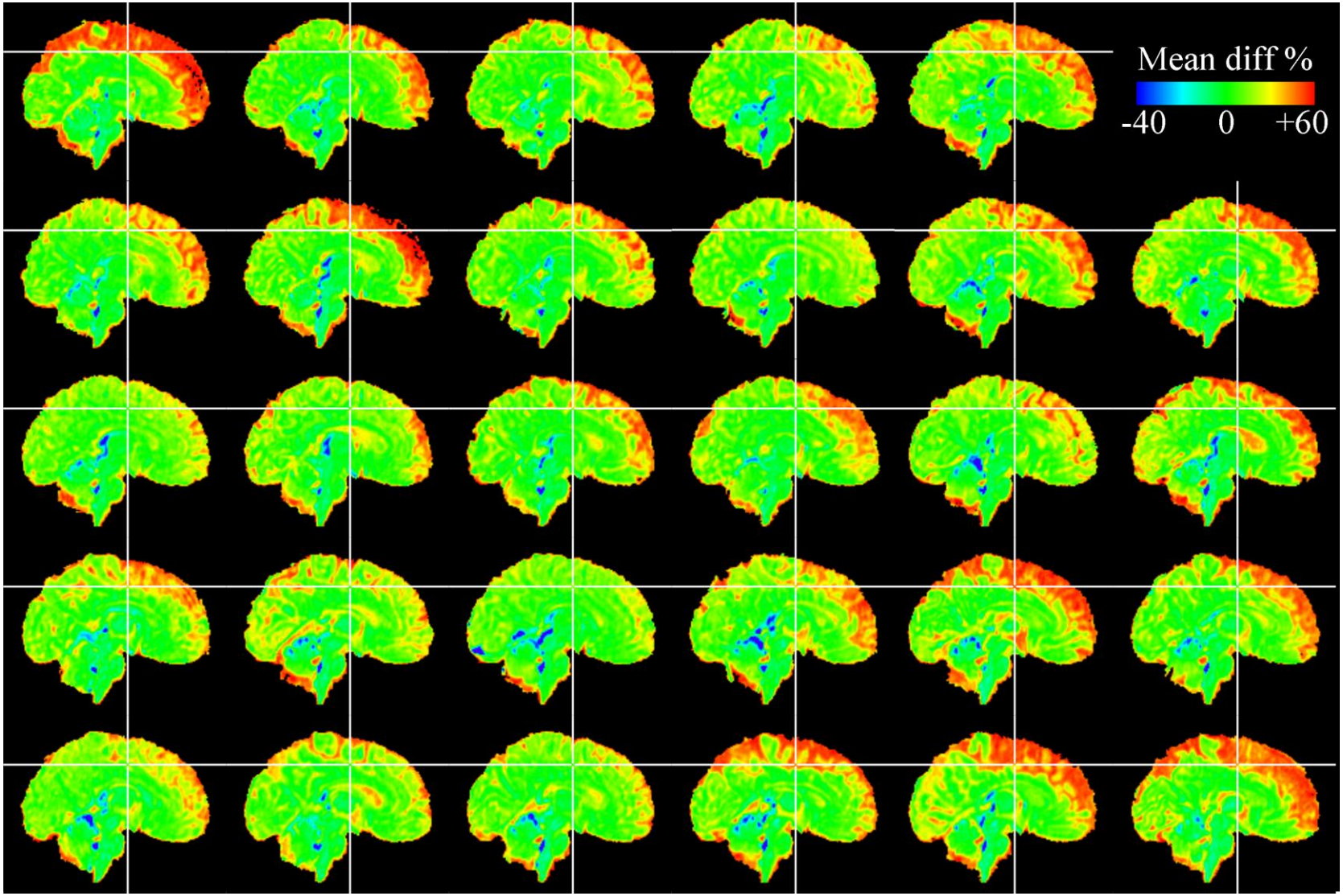


Figure 64. Preterm infant mean distance images These show how far (in percent) each preterm infant voxel was from the mean normal atlas. The red areas indicate hypointensities (reduced grey matter/ increased cerebrospinal fluid proportion) in preterms that are generally not found in full term infants.

The reference used to derive the images in Figure 64 was based on the mean intensity of 36 normal infant subjects. I have already discussed (chapters 3 and 4) and demonstrated (chapters 6, 7, 8) the limitations of mean values of brain structure and they still apply here. However, when we have collected a sufficient amount of normal infant data ourselves, I will create a percentile rank template to determine if this can improve sensitivity to detect pathology associated with preterm infancy.

While I did not have sufficient data to create a normal percentile rank atlas for infants, I did have sufficient data to create percentile rank atlases for adults. I describe these in the following sections.

9.4.2 Percentile rank "adultspan" BRAINS atlas

As the review in chapter 3 showed, there was a lack of brain MRI atlases for studies across adulthood. The few existing atlases were created with less than 100 subjects each and were based on mean voxel values. I therefore created a percentile rank "adultspan" atlas with 160 subjects aged 25–90 years.

9.4.2.1 Methods

To derive the standard space for the adultspan (25–90 years) atlas I followed the procedure described in chapter 5. Briefly, each subject was 6 pt registered to MNI space and a mean intensity image calculated; each subject was then 12 pt registered to the 6 pt mean intensity image and a final mean intensity image calculated. The 12 pt mean intensity images from the MNI152 atlas, my adultspan (25–90 years) and aged (60–90 years) atlases are shown in Figure 65. This shows that an atlas representative of the span of adulthood has ventricles and sulcal spaces larger than the MNI152 atlas but not as large as an aged atlas. Further, my adultspan atlas is more representative of the size of the average adult brain (because I removed residual variance that led to the overestimated size of the MNI atlas; chapter 3).

The MRI intensity range in each subject was standardised to the intensity range of the adultspan standard space using the SIS procedure described in chapter 5. The 2.5^{th}, 25^{th}, 50^{th}, 75^{th}, and 97.5^{th} percentile ranks of intensities were then calculated in each voxel (with the procedure described in section 8.2.2.2.2). The result of which was a four dimensional (4D) percentile rank atlas of the normal levels and limits of voxel intensities for subjects 25–90 years.

9.4.2.2 Results

The standard space for the adultspan atlas is shown in Figure 65 alongside the standard space for the aged adult atlas and the MNI152 atlas. The results from SIS are shown in Figure 66, illustrating that all subjects, from 62 different sites (ADNI, OASIS, and Edinburgh), were in the same intensity range (this procedure was validated in section 5.4.3). The subsequently derived percentile rank adultspan atlas is shown in Figure 67. Although data were not available at the time of writing, I will use this atlas to investigate the pathology associated with schizophrenic individuals (that may be too subtle to detect with parametric methods; Shenton et al., 2001; Job et al., 2002). Further, by ranking subjects with this template, data in subsequent lifecourse studies (25–90 years) will be more amenable to nonparametric analyses, e.g. the Wilcoxon–Mann–Whitney test, while still in a relative transformation space, i.e. rather than a Box–Cox (Box and Cox, 1964) or logarithm transform, their voxel values are replaced with percentile ranks of "how normal" they are. Voxel–based brain ranking was fully described in chapter 8.

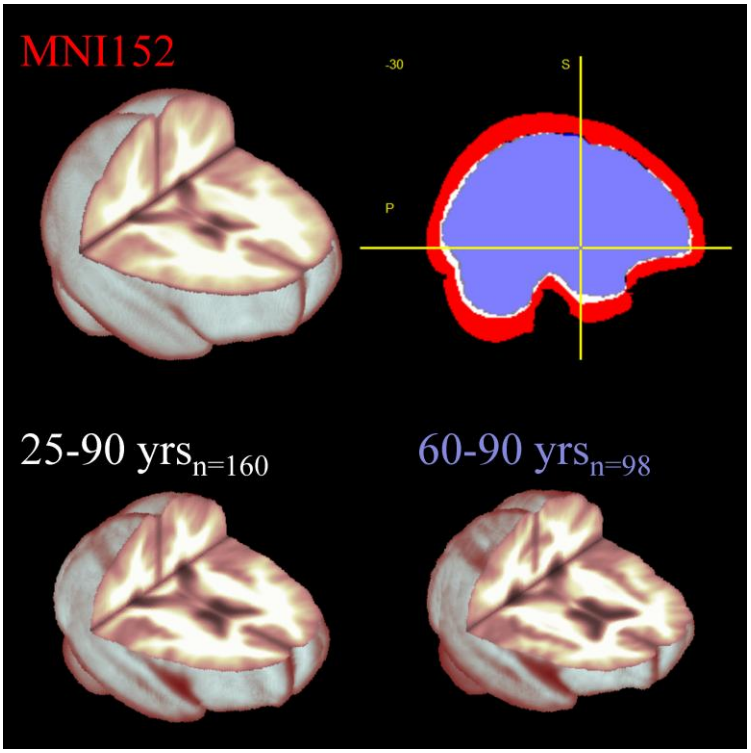


Figure 65. MNI152, adultspan, and aged atlases
The sagittal view of the atlases overlaid with one another is shown in the top right; this illustrates the differences in brain sizes between ages, e.g. brain size generally shrinks with age. MNI152 is shown in red, "adultspan" in white, and the aged atlas in blue.

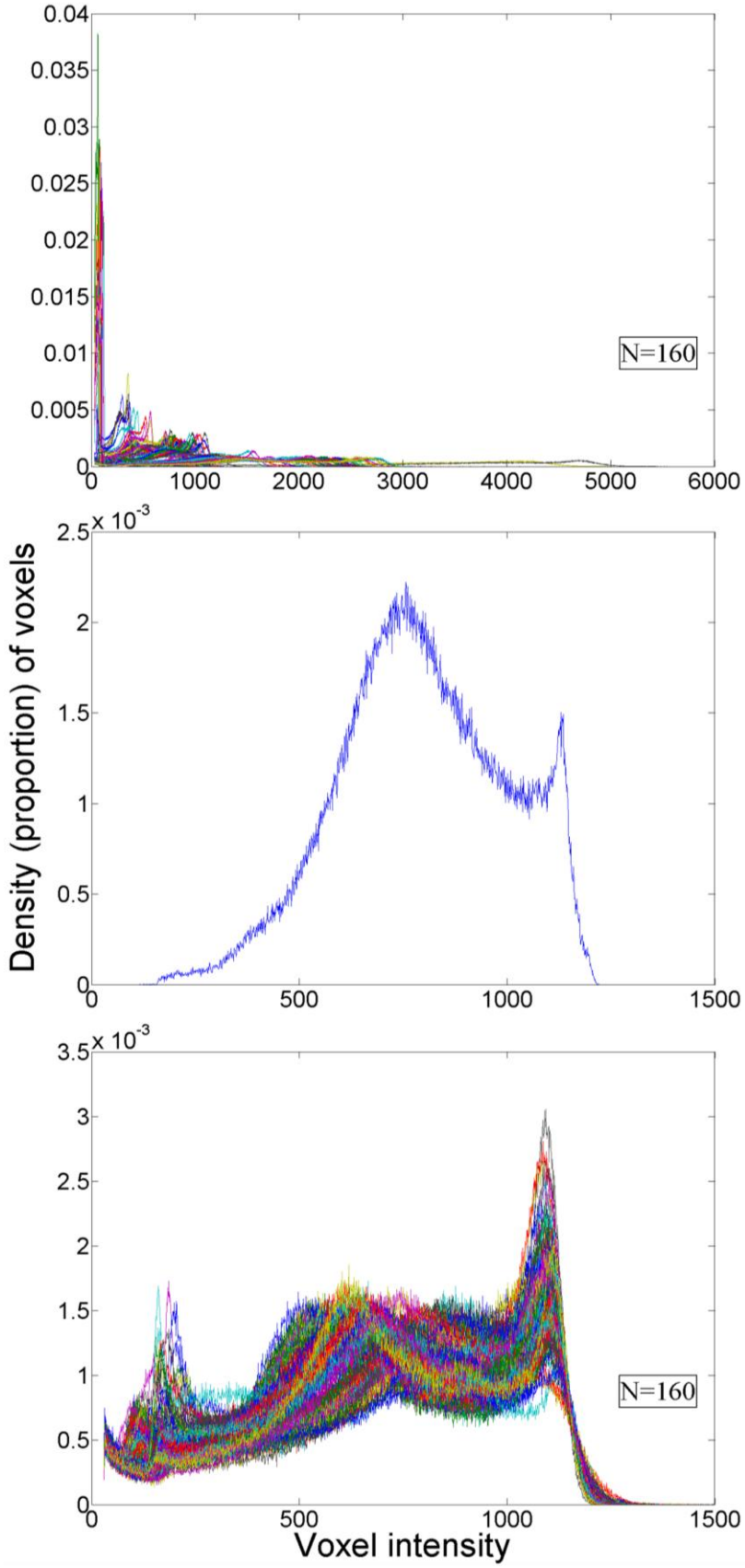


Figure 66. SIS in the “adultspan" BRAINS atlas subjects
The original voxel intensities of 160 different subjects (25–90 years) from 62 different sites (ADNI, OASIS, and Edinburgh; top panel) were standardised (bottom panel) to the standard space intensities (middle panel) using the procedure described in chapter 8.

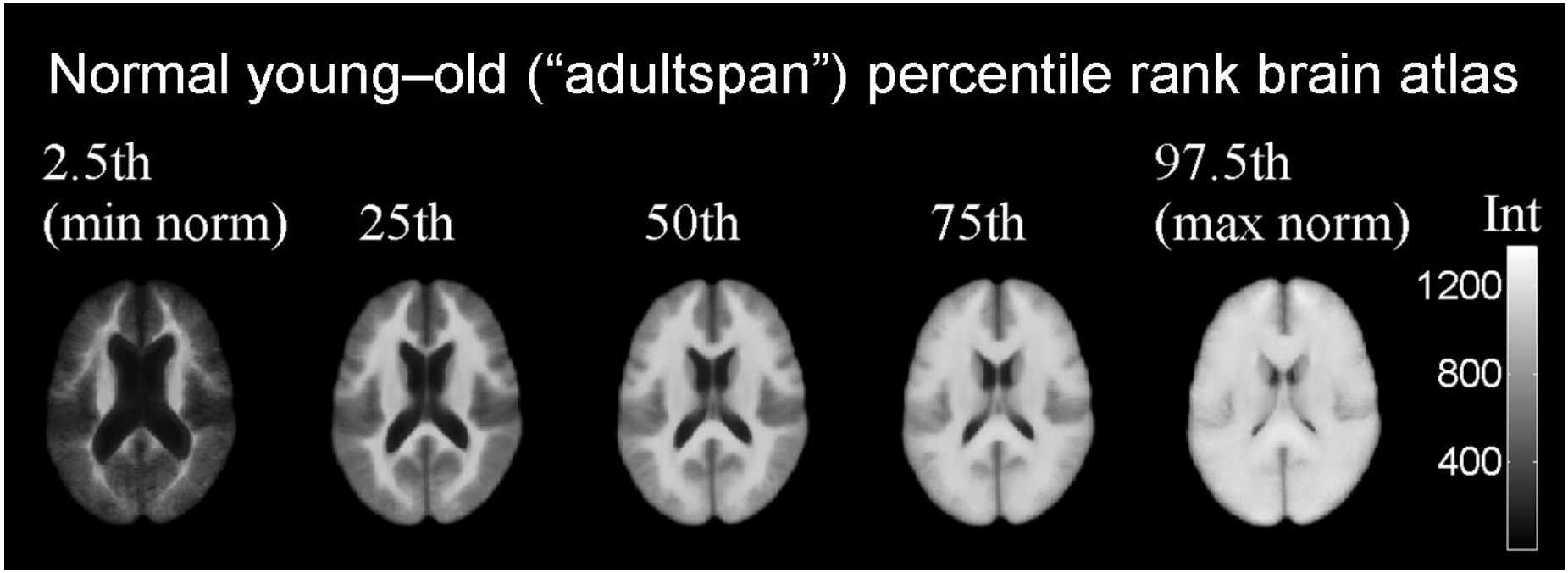


Figure 67. “Adultspan” (25–90 years) whole brain percentile rank atlas

This shows the normal range of intensities (from the 2.5th to 97.5th percentile rank values) in each voxel for subjects aged 25–90 years; min=minimum; max=maximum; norm=normal value; Int=voxel intensity. Because it is calculated at the voxel level these images do not literally correspond to, for example, a single brain with lots of cortical cerebrospinal fluid/grey matter (at the 2.5th percentile). Rather, for the 2.5th percentile, these are just the minimum normal values at each voxel shown (in any given brain image, CSF has the lowest, GM the medium, and white matter the highest intensity values).

9.4.3 Percentile rank "aged" BRAINS atlas and AD pathology map

I created a percentile rank "aged" whole brain atlas using SIS and 236 normal subjects aged 60–90 years because of the issues I identified with GM proportion images and atlases in aged subjects (chapter 5).

9.4.3.1 Methods

The percentile rank aged whole brain atlas was created using the same procedure as the adultspan atlas. Two hundred and thirty–six normal subjects from ADNI and OASIS (aged 60–90 years; described fully in chapter 5) were registered to the aged standard space (section 5.3.5.1) with Nsurf registration (section 5.3.6). I then standardised the voxel intensity range of each subject to the standard space voxel intensity range using SIS (section 5.4.3). After which I calculated the 2.5^{th}, 25^{th}, 50^{th}, 75^{th}, and 97.5^{th} percentile ranks of intensities in each voxel to produce a four dimensional percentile rank atlas of the normal levels and limits of voxel intensities for subjects 25–90 years.

I then applied the same processing steps to 224 AD subjects from ADNI and OASIS (aged 60–90 years) that were described in chapter 5. This allowed me to classify each of their voxel values with the percentile rank atlas.

9.4.3.2 Results

The percentile rank aged whole brain atlas is shown in Figure 68. By ranking 224 AD subjects with this atlas I produced a "pathology map", i.e. the median of the ranked images where voxels less than or equal to the lower quartile of normal values were marked in red. The AD subjects had mixed CDR greater than 0, i.e. 0.5-1. This highlighted established regions of pathology associated with AD, e.g. ventricular enlargement and MTL and hippocampal atrophy (Figure 69). It also highlighted white matter hypointensities (hyperintensities on T2) that are emerging as a potential marker of cognitive decline and dementia (Breteler et al., 1994). Further, it highlighted GM regions that I found parametric tests to produce false positive results, and a potentially new marker of Sylvian fissure (lateral sulcus) enlargement. From qualitative assessment of the pathology map, the Sylvian fissure showed a large area of voxels that were less than the lower quartile of normal values. This is consistent with the

large areas of lower quartile values in established regions of pathology, e.g. lateral ventricles and hippocampus. Although not a definitive new marker, the proximity of the Sylvian fissure to the medial temporal lobe, and similarities of ranks between the Sylvian fissure and established pathology areas, suggests this area merits further investigation with more subjects and nonparametric analysis.

The previously identified markers were found using a variety of techniques and sequences whereas I replicated them with an atlas–based ranking approach in a single sequence. This suggests that the SIS percentile rank atlas may highlight previously unknown pathology of other neurological disorders, e.g. schizophrenia. However, these findings are by no way definitive and are only meant to drive hypotheses for further analyses. They may also support individual diagnoses, e.g. by *a priori* highlighting specific regions for clinicians to investigate in new individuals suspected of developing cognitive decline.

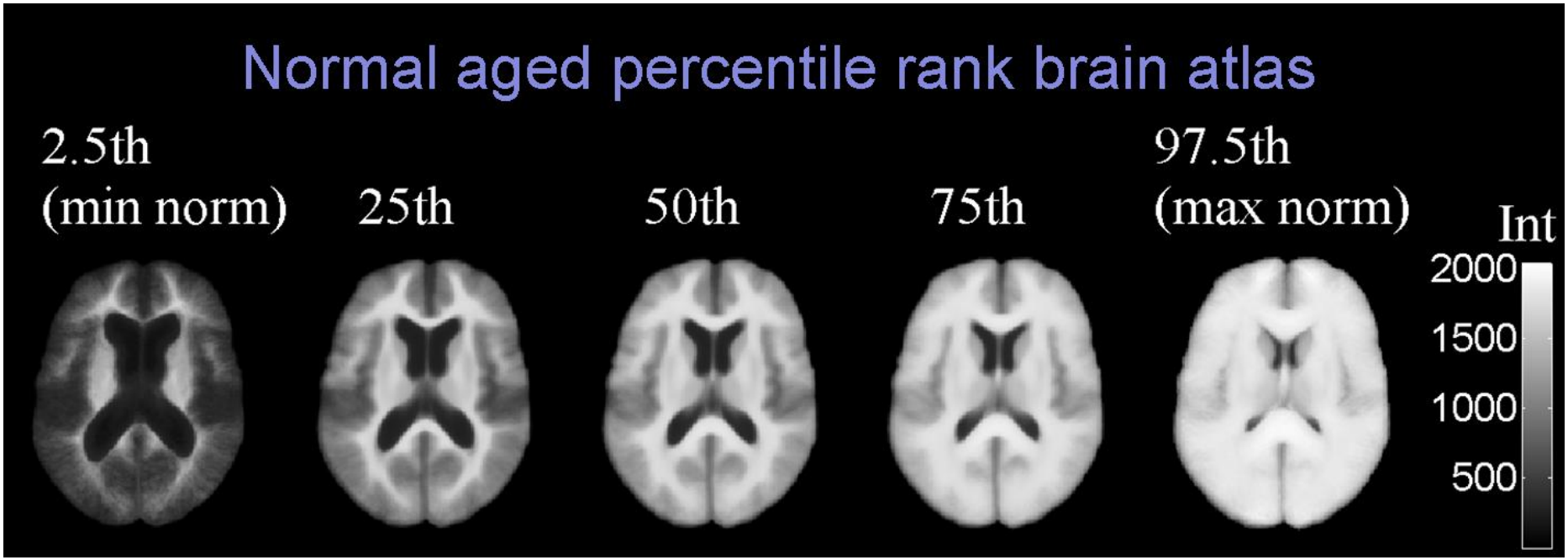


Figure 68. Aged adult (60–90 years) whole brain percentile rank atlas

This shows the normal range of intensities (from the 2.5th to 97.5th percentile rank values) in each voxel for subjects aged 60–90 years; min=minimum; max=maximum; norm=normal value; Int=voxel intensity. The inclusion of only subjects aged ≥60 years led to a median (50th percentile) image with slightly larger lateral ventricles and sulcal spaces than the "adultspan" atlas (figure 67). Because it is calculated at the voxel level these images do not literally correspond to, for example, a single brain that is made up of almost entirely white matter (at the 97.5th percentile). Rather, for the 97.5th percentile, these are just the maximum normal values at each voxel shown.

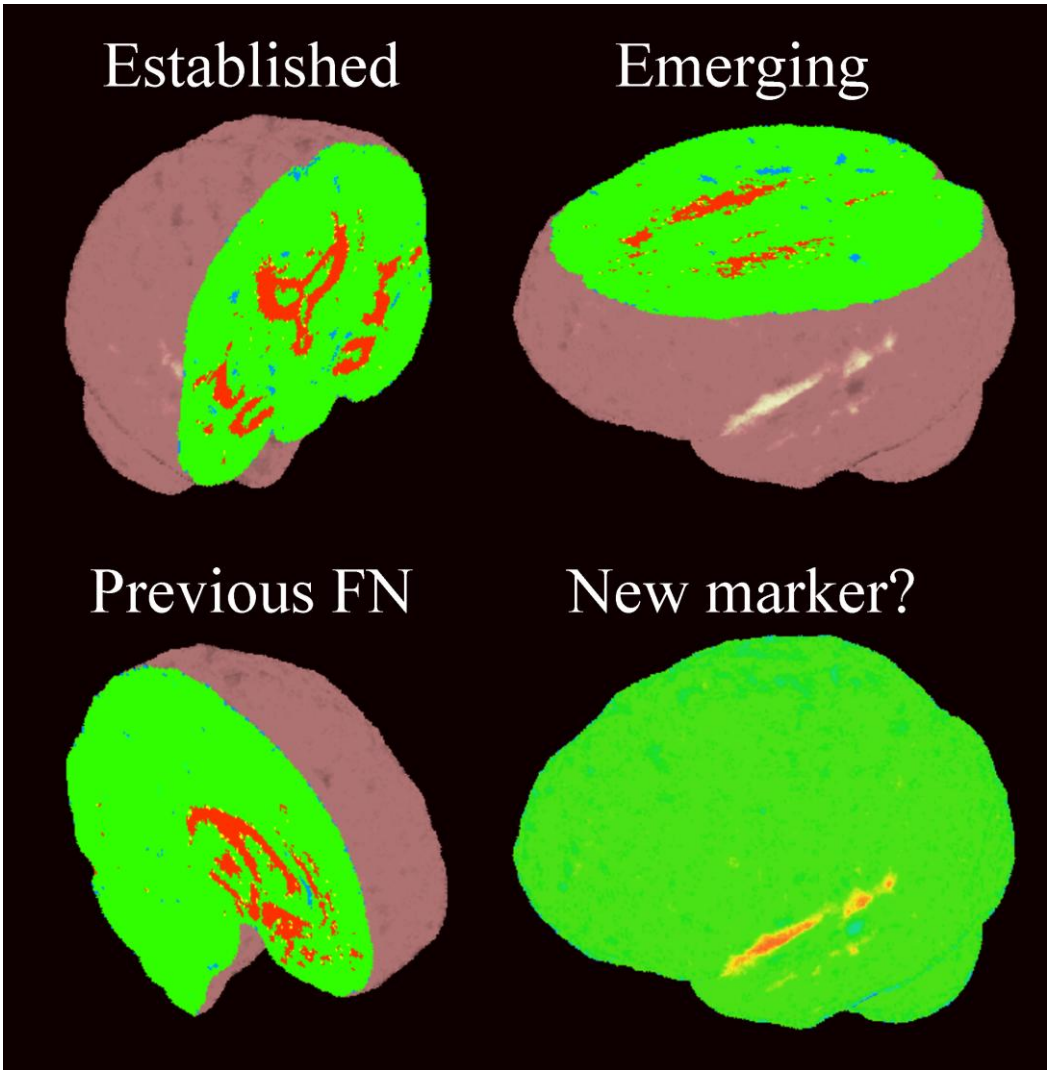


Figure 69. Alzheimer's disease pathology map
This is the median of 224 AD subjects (aged 60–90 years) from ADNI and OASIS that were ranked with the aged percentile rank brain atlas. This highlights regions in AD where voxel intensities are at the lowest quartile of normal values (or below) and illustrates pathology established in the literature, medial temporal lobe atrophy and ventricle enlargements (top left), potentially emerging pathology, white mater hypointensities (hyperintense on T2), areas where parametric methods previously produced false negatives (FN), in the frontal region, and a potentially new marker, Sylvian fissure (lateral sulcus) enlargement.

9.5 Discussion

In this thesis I have attempted to determine the amount of brain imaging and metadata required to effectively describe the full range of normal ageing brain structure. The large amounts of data that are required to understand normal and pathological brain ageing, and the current lack of these data, led us to develop the BRAINS bank. Many brain image banks are currently available but their representation of normal older people is limited (chapter 2). Medical, cognitive, and image sequence data for subjects

currently available on the internet may not be sufficient to create robust statistical models of normal ageing brain structure for identification and prediction of pathology. I discuss here how BRAINS may provide the data and methods needed to do this.

9.5.1 BRAINS image bank

The BRAINS image bank, in the first instance, will consist of 1595 different normal subjects aged <1-80 years (~50% female). These were acquired in 10 separate studies that were conducted at different points in time. While this may at first seem a limitation, e.g. due to technology updates, it is actually a strength. Data are often collected at different points in time (e.g. in longitudinal studies) and with the data that we have, I may identify the potentially spurious effect of year of scan between individuals (there is clearly a strong and significant effect across ages). For example, in a group of subjects of the same age, I could include the year they were scanned, e.g. 2005 vs. 2010, as a controlling variable. Adding this effect to longitudinal data may then control for factors such as technological and image sequence advances. Other factors that may confound studies of normal ageing include blood pressure and prescription medication use (Mazziotta et al., 2009). These were not always provided in existing banks but were provided in BRAINS when collected in the original studies.

We collected a vast array of medical and cognitive data for the aged BRAINS subjects. Unlike in previous banks (Mazziotta et al., 2009), these data were not used to exclude subjects but provided for others to make that exclusion. Our normal subjects therefore had no dementia, other debilitating disorder or history but common ageing indications such as hypertension. This means that they are generally not "super-normal" but investigators can exclude subjects on these variables if appropriate for their study. Following medical testing each subject received a battery of medical tests and, as long as they were within normal limits (e.g. MMSE>27), we did not exclude or stratify subjects based on these measures. Again, these data will be provided for individuals to exclude/ co-vary as they see fit. After we determined that subjects were normal via these medical and cognitive data, they underwent a series of imaging sequences.

All of the aged (>60 years) subjects in BRAINS were acquired in multiple imaging sequences. The T1-weighted images provide the basic data required to

segment whole brain and subregional volumes. This is not always sufficient for aged subjects, e.g. misclassifications of WMH in chapter 5. The T2, T2*, and FLAIR sequences provided in BRAINS will support robust segmentation of these volumes (Wardlaw et al., 2011; Aribisala et al., 2012). We have also acquired longitudinal data for the majority of the aged subjects. This will allow us to compare the cross-sectional models described here with future longitudinal models. Further, these data will allow us to test whether my atlases can support diagnoses/ predictions of cognitive decline and neurodegenerative disease.

9.5.2 BRAINS atlases

BRAINS atlases and voxel-based pathology detecting methods were developed for subjects from <1 year to 90 years. Normal infant (<1 year) subjects were not available at the time of this work and so I used an existing mean based atlas (Fonov et al., 2009) to illustrate how I may automatically detect pathology in preterm infants.
This allowed me to determine how far away preterms were from the general pattern of normal infant brain structure but did not allow me to determine whether these differences were out with the normal range. Although the vast majority of subjects showed large (>50%) deviations from normal infancy values in the interhemispheric fissure and inferior cerebellum, some showed smaller (~10-20%) deviations.

Prior to this atlas-based assessment I did not perform bias field correction, but I did normalise all subject intensity ranges using SIS (described in chapter 5). Bias field appeared to unduly alter the deep GM structures and there are several challenges to reliably segmenting tissues from infant brains, e.g. the grey/white boundary is inversed on T1–weighted images in infancy. Future work will address these challenges and whether or not there are any ill effects in intensity standardisation that cannot be determined visually. Furthermore, a future percentile rank atlas will show whether or not the deviations I highlighted are within normal limits.

I was able to develop percentile rank atlases across the span of adulthood (25-90 years) and old age (60-90 years). These atlases were derived in the coordinate framework of the widely used Talairach/ MNI space, but did not inherit the limitations I discussed previously (chapter 3). For example, neither of my atlases were volumetrically symmetrical and had overall brain size close to that of most people. My atlases may therefore be more amenable to image processing (by being similar to

more people's brain size they give registration algorithms a better chance of good correspondence) and subsequently reduce potential errors in atlas-based assessment, e.g. those due to misalignment of overall brain size.

Atlas-based assessment with the aged adult atlas allowed me to create AD pathology maps. That is, from the median classification image of 224 AD subjects, I highlighted regions that were equal to or below the lowest quadrant of normal values. Using just their normalised T1-weighted images, this map highlighted much of the pathology identified in literature that used a number of sequences, e.g. hippocampus atrophy and WM hypointensities (hyperintense on T2). Further, previously identified regions of false negatives in GM images (chapter 8) were highlighted and a potentially new marker was found in the Sylvian fissure (lateral sulcus). Although temporal lobe atrophy and WMH markers are supported by a large body of evidence (Breteler et al., 1994; Jack Jr et al., 1997; Thompson et al., 2003), there is much less independent evidence to support frontal lobe and Sylvian fissure expansion as markers of AD. Therefore, my pathology maps are not intended to be definitive measures of brain structure in diseases like AD. Rather, they are intended to drive hypotheses for further analyses and support individual subject diagnoses. For example, a future study may specifically investigate the potential WMH marker in white matter tracts or measure cortical thickness in the Sylvian fissure. In clinical diagnoses, a pathology map may provide a reference for areas to check in subjects that are suspected to have incipient disease. Many more subjects will be required to create robust and representative atlases but I have at least described a method for doing this when these subjects become available.

While most commonly used atlases were representative of only one race or nationality, my "adultspan" atlas was created with normal subjects from the United Kingdom and United States. This suggests that it may be useful for increasingly international brain imaging projects (Mueller et al., 2005; Ellis et al., 2010; Simmons et al., 2011). The next section discusses how BRAINS fits into the context of existing international banks and atlases.

9.5.3 The place of BRAINS in the context of existing brain image banks and atlases

Banks of brain data have been in development since at least the 1980s (Foulkes et al., 1988) and image banks started to become prominent in the 1990s (Cocosco et al.,

1997). Prominent brain image centres, such as the MNI and LONI, developed their strong international reputations partly on the back of their databank and atlas work (Evans et al., 1993; Toga, 2002). While they have successfully addressed many of the concerns with historical brain atlases and the previous lack of validation data, my visits to these centres also demonstrated there are still gaps that may be filled by BRAINS.

The ADNI databank is housed at LONI and is an extremely valuable neuroimaging data source. Much of the present work was partly based on analyses of ADNI data. It was designed to provide earlier diagnoses for AD and so, in a range of MR sequences and with rich cognitive and medical data, recruited over 800 AD and normal control subjects. There are, however, some limitations to ADNI. The youngest of the control subjects are aged 70 years. People as young as forty may develop AD. Prodromal pathology may occur in the brain up to 10 years or more prior to clinical symptom onset (Tondelli et al., 2012). This suggests that normal subjects younger than 70 years are required for effective earlier diagnoses. These, and subjects from across the lifespan, will be provided in BRAINS. This means that BRAINS will be useful for investigating neurological disorders across the lifespan, from autism to Alzheimer's.

Further to BRAINS data, my atlas method also represents a novel approach. The most widely used brain image atlases were developed at MNI (Evans et al., 1993; Ashburner and Friston, 2000). The so-called "MNI" or "ICBM" atlases/ templates, they were derived from the mean voxel values of several hundred subjects aged 18-44 years. These are very useful for image processing, e.g. tissue segmentation, voxel-based reporting. However, they may introduce bias into studies of subjects aged younger than 18 and older than 40 years, e.g. overexpansion of aged brains (Buckner et al., 2004). Moreover, these mean-based atlases may lead to errors if used for voxel-based classification (chapter 8). BRAINS atlases will be derived from subjects across the lifespan using robust percentile rank voxel values. The next and final chapter discusses the implications, limitations, and contribution of these methods.

10. Discussion and conclusion

10.1 Contribution

Robust methods are required to understand the complexity of the human brain and how it changes through life and disease. Following assessment of existing methods, I provided novel methods for highlighting changes that occur in the brain before and during neurological diseases. Here I discuss the conclusions from the hypotheses set in Chapter 1 and the main contributions of this work:

A1. Collated and summarised previously disparate information about brain image data publicly available worldwide

A1.1 Described how many existing brain image atlases were built and what they may be used for.

A1.2 Reviewed the most commonly applied statistical models in brain imaging.

H1. I rejected the hypothesis that parametric atlases adequately describe the distribution of brain structure across the lifecourse:

A2. I therefore built novel percentile rank brain image atlases of the "adultspan" (25-90 years) and ageing (60-90 years)

A2.1 Highlighted the potential for errors in individual assessments with current atlas methods.

A2.2 Provided a more robust method for assessing brain structure in individuals.

A2.3 Developed "atlas-based clinical rank transformation" to prepare data for nonparametric testing while maintaining a relative frame of reference.

A2.4 Determined that a robust atlas of the distribution of brain structure across the lifespan would require approximately 6,370 subjects.

H2. I rejected the hypothesis that parametric statistical models adequately describe the distributions and boundaries of brain structure in normal ageing and disease:

A3. I therefore implemented novel percentile rank statistical models of brain structure

A3.1 Highlighted the potential for errors in definition of the distribution of ageing brain volumes with current (mean-based) methods.

A3.2 Provided a method for directly calculating distribution levels and boundaries (percentile ranks) of ageing brain volumes.

10.2 MRI brain structure data

A systematic review of structural MRI brain databanks found that there are fewer than 350 cognitively normal older subjects (≥60 years) publicly available worldwide and 98 of these are openly available worldwide. I found many other subjects were available but they did not have cognitive and medical data to support their definition of normal. The definition of normal subjects should be supported with thorough cognitive and medical testing to show that they have no cognitive or other debilitating disorder but clinical characteristics that are common in ageing e.g. prescription medications, diabetes, hypertension and arthritis (Jernigan et al., 2001; DeCarli et al., 2005; Grady et al., 2006; Marcus et al., 2007c; Ellis et al., 2009; Farrell et al., 2009; Mazziotta et al., 2009; Marcus et al., 2010). As long as these characteristics are carefully documented and searchable, this will allow others to draw subjects appropriate for their study group and/or representative of the wider population. To fully represent the range of brain structure seen in ageing, multiple image sequences, e.g. T1, T2, T2*, FLAIR, are required. Related to this, these sequences are also required to support reliable image processing across the lifecourse. For example, to reliably segment CSF and white matter hyperintensities, T2 and FLAIR sequences are required (Wardlaw et al., 2011; Aribisala et al., 2012). The wide range of brain structure seen in older ages requires that these data are available for all subjects, otherwise this variance may not be adequately quantified, i.e. by excluding subjects without the required images I may artificially reduce the natural range of variance.

To reasonably define the range of brain structure in a group of subjects I determined that approximately 70 subjects would be required (percentile ranks in Figure 49 converged to within 5% after 90 subjects). This suggests that a robust atlas of the distribution of brain structure across the lifespan (0-90 years) would require approximately 6,370 subjects (70 subjects per year × the 91 years of life included).

10.3 Brain structure atlases

A formal review of existing MRI brain structure atlases (chapter 3) found that the vast majority were derived from the mean intensity images of young to middle aged adult subjects. These are appropriate for studies of these subjects but may introduce bias in

studies of other ages, e.g. older subjects (≥60 years; Buckner et al., 2004). It has therefore been recommended that atlases should represent the range of subjects under study (Buckner et al., 2004; Rorden et al., 2012). For adult lifecourse studies, i.e. 18–90 years, there were two existing atlases available but these only had 40 subjects at most (Desikan et al., 2006; Rohlfing et al., 2010). Further, mean intensity atlases may not well represent the other levels and limits of brain structure. I therefore created larger, percentile rank atlases specifically for studies of ageing subjects and adult lifecourse studies.

I created parametric (mean–based) and nonparametric atlases of the range of GM proportion in normal ageing (60–90 years). The parametric atlas produced nonsensical lower limits of GM proportion in voxels throughout the brain, i.e. proportion limits less than zero. This led to false negative (normal) classifications in the mid caudal region in two samples of AD subjects. Parametric models assume that the lower clinical limit (2.5^{th} percentile rank) is equal to the mean minus two SD. However, in many voxels throughout the brain, specifically in the MTL and hippocampus, the parametric limit was less than the 2.5^{th} percentile rank. This led to a number of false positive (abnormal) classifications in the MTL and hippocampus in the two samples of AD subjects. I therefore recommend that, when creating an atlas to describe the levels and limits of brain structure, percentile ranks are directly calculated using a nonparametric (rank–based) method rather than a mean–based method.

I created nonparametric whole brain atlases for the normal range of brain structure in ageing (≥60 years) and across the adult lifecourse (25–90 years). These were created with intensity normalised T1 images rather than tissue proportion images that may include misclassified voxels, e.g. white matter hypointensities (hyperintensities on T2) as GM. By ranking AD subjects with the aged adult atlas I was able to produce a "pathology map" that automatically highlighted established areas of pathology in AD, e.g. MTL atrophy, while also highlighting potentially novel markers of the disease, e.g. Sylvian fissure atrophy. This map is not definitive but may form hypothesis for future studies and support diagnoses of individual patients.

Although I did not have sufficient data to create a percentile rank template of normal infant brain structure I calculated "mean differences" from an existing mean intensity normal infant template. This automatically highlighted regions of pathology, e.g. increased interhemispheric fissure CSF, found in preterm infants by an

experienced neonatologist. This suggests that a future percentile rank atlas may support diagnosis of abnormal brain structure in preterm infants. Whole brain templates were created for infants because of the difficulty in segmenting tissues at this age (WM/ GM contrast is low and reversed from adulthood). However, despite the potential for misclassifications, tissue image atlases are easier to create and useful for voxel–based ranking in older subjects.

Voxel–based GM ranking with the nonparametric atlas allowed assessment of individual subjects without the need to apply a statistical test, e.g. *t*–test, that may not be appropriate for individuals (Scarpazza et al., 2013). This method highlighted the subtle brain structure changes associated with AD in individual subjects. Further, by ranking individual subjects with the nonparametric atlas it transformed data to be more amenable for subsequent nonparametric statistical testing while maintaining a relevant transformation space, i.e. "how normal" each voxel value was. This transformed the shape and variance of individuals into a common framework which is a requirement for nonparametric tests such as the Wilcoxon–Mann–Whitney test. However, as with any transformation, there are inherent limitations to this (discussed in section 8.3.4.2) and therefore should only be applied with due consideration. For situations where data transformations are not desirable, I developed and/or implemented a number of novel statistical modelling methods.

10.4 Statistical models of brain structure across the lifecourse

A formal review of existing statistical models of normal brain structure across the lifecourse found that the vast majority of results were obtained with parametric methods. These methods are successful at defining general (mean) changes in brain structure but, given their varying robustness under departures from their assumptions, may not be the most appropriate method for defining changes at the limits of normality (Elveback et al., 1970; Meehl, 1978; Ashburner and Friston, 2000; Courchesne et al., 2000; Good et al., 2001; Freedman et al., 2007). I therefore developed nonparametric statistical models that may be more appropriate for defining changes in brain structure at levels and limits other than the mean.

It was known that whole brain tissue volume generally decreases with age, but it was not known whether general (mean) decreases were reflected at percentile ranks

above and below the mean, e.g. clinical limits. I therefore developed percentile rank models of whole brain tissue volumes in normal ageing and AD and compared these to the mean differences. I found that mean differences in whole brain tissue were not replicated at other percentile ranks in the publicly available data (ADNI and OASIS databanks). This suggests that to accurately define the levels and limits of normal ageing whole brain tissue volumes, nonparametric percentile rank models will be required.

As neurological diseases are associated with subtle changes that may occur earlier in life (Shenton et al., 2001; Tondelli et al., 2012), I also assessed whether mean differences in regional brain volumes between younger adults were representative of percentile rank differences. In a sample of 80 young to middle adult (25–60 years) subjects, I found that mean differences were not representative of nonparametric percentile rank differences. This was particularly evident in the hippocampal region and caudate. The hippocampal region has been implicated in several neurological diseases, e.g. AD and schizophrenia (Jack Jr et al., 1997; Job et al., 2002), therefore it is important to accurately define the normal levels and limits of volumes in this region. Although a sample of 80 subjects is very limited, these preliminary data suggest that nonparametric percentile rank, rather than parametric, models may be required for accurate definition of these distributions and boundaries.

10.5 Benefits and limitations of parametric and nonparametric methods

This work reviewed existing methods for assessing brain structure, many of which were based on parametric statistical models. These are strong, well validated methods for making general (mean) statistical inferences (Freedman et al., 2007; Hogg and Tanis, 2010). The point of this work was not to condemn these methods but to demonstrate that when their prerequisites are not met, they may lead to unexpected and potentially unreliable results. This is especially true when attempting to define the boundaries of normal and pathological brain volumes (chapters 6 [whole brain], 7 [regional], and 8 [voxel]). Although this may appear obvious, I found little in the way of previous work that assessed the effect of statistical assumptions in structural brain imaging. Further, I found few reports that explicitly stated they checked the assumptions of their models (chapter 4). Smoothed brain structure has previously

been shown to be at least approximately Gaussian (Rorden et al., 2007) and it may be that in other reports the use of parametric models in itself indicated that their assumptions were met.

My aim is therefore not to criticise others, but to show that there are many other statistical methods that may be more useful when parametric assumptions are not met. In particular, as noted by Sir Francis Galton in 1889 (chapter 1), there can be many interesting statistics other than those based on means. For example, I showed in chapter 6 that percentile rank may indicate how one person will change compared to another at the opposite extreme of normality, e.g. subjects at the lowest percentile ranks of brain tissue may atrophy quicker than those at percentile ranks above the mean. This, however, is the first structural brain imaging implementation of percentile rank regression and further testing will be required to determine if it is actually true. There are therefore a number of benefits and limitations of parametric and nonparametric methods. These are summarised in Table 46.

Table 46. Benefits and limitations of parametric and nonparametric methods

Parametric		Nonparametric	
Benefits	*Limitations*	*Benefits*	*Limitations*
Robust and reproducible method when data are equally Gaussian.	Provides unpredictable results when data are not equally Gaussian, e.g., may over or underestimate the central tendency of data.	Provides robust measures for the full range of percentile ranks with a sufficiently large sample.	Cannot be used to describe the full range of percentile ranks with a sample less than 100 subjects.
Provides easily interpretable results when data are equally Gaussian.	Results may not be extrapolated to values beyond the mean if data not equally Gaussian.	Provides additional information for differentiating groups, e.g., whether mean differences are caused by skewed distributions.	Produces, at least in the work presented here, 4 additional statistics (percentile ranks further to the central tendency) that may be difficult to interpret.
Has increased power to detect subtle mean differences when data are equally Gaussian.	May overestimate mean differences when data are not equally Gaussian.	Provides a “true” measure of the central tendency between groups.	The “true” differences are only true in the samples assessed. These differences are not generalizable to the population without repeated testing in large samples.

There are further limitations in my work, related to image quality (MRI field strength), image processing, the representativeness of atlases, and validation of results (number of subjects). These are discussed in the next section.

10.6 Limitations

The methods that I developed in this work are limited by a number of factors related to image quality (MRI field strength), image processing, the representativeness of atlases, and validation of results.

10.6.1 Image quality (MRI field strength)

The brain MRI used in this work were acquired at 1.5 T. Substructures such as the hippocampus are visible at 1.5 T however subfields of these structures are not clearly visible. This means that subtle differences between groups of controls and neurologically diseased groups, e.g., Alzheimer's disease, may not be apparent at 1.5 T, and therefore not apparent in the present study. Higher field strengths, e.g. 3 to 7 T, will be needed to clearly differentiate structures such as the boundary between the hippocampus and amygdala within the hippocampal complex (Wisse et al., 2014). There are a broad range of hippocampal sizes in normal ageing and AD (Jack Jr et al., 1997; Ferguson et al., 2010). This means that longitudinal imaging data are required to differentiate the scarring of disease and markers of disease. The cross-sectional data used here is limited in differentiating disease scarring from markers of disease. For example, it may have been that some AD subjects always had smaller hippocampi and it is the relative change, rather than the absolute size, that is a marker for disease. Future work with longitudinal data and higher field strengths will be required to determine subtle disease markers prior to onset.

Subtle longitudinal changes within individuals that occur through time may be obscured at 1.5 T (Shenton et al., 2001). Furthermore, there is variability within individuals between scanning sessions at different times of the same day and between different scanners and sequence parameters (Gountouna et al., 2010). Future work with data from the same subjects across different scanners and varying time points, e.g. within a week and across years, will be required to determine whether parametric

or nonparametric methods are better at differentiating spurious changes in MRI signal from signs of disease.

The software I developed is not specific to 1.5 T, scanner, nor sequence parameters and may be implemented with any field strength, scanner, and sequence parameters. This will allow for an assessment of the impact of field strength, scanner, and sequence parameters in nonparametric versus parametric assessments of brain MRI.

10.6.2 Image processing

This work attempted to define normal and pathological variation in brain structure across the lifecourse. For that reason I implemented very specific image processing methods. In my voxel–based statistical analyses, I developed a novel registration (spatial normalisation) method, Nsurf as described in chapter 5. Nsurf attempted to address issues with conventional registration methods. Linear, i.e. 12pt affine, registration does not always adequately account for head size. Nonlinear registration removes a large part of the variance of interest, e.g. lateral ventricle volume. Nsurf performed nonlinear registration on mask (binary) images so to adequately account for head size but retain within brain variance. This means that my results are not directly comparable with previous studies. Additionally, unlike many previous studies (Ashburner and Friston, 2000; Good et al., 2001), I did not perform voxel smoothing. I did not do this because, by essentially averaging adjoining voxels into one, it may have obscured subtle differences between normality and pathology. Further, there are not yet consistent, quantitative means for determining smoothing parameters (Zhao et al., 2012). Furthermore, one reason for smoothing is to make data more Gaussian distributed (Ashburner and Friston, 2000; Good et al., 2001); this was not required for my nonparametric methods. But despite my reasons for not smoothing, this may have led to a degree of noise in my results and may partially explain the non-Gaussian voxel distributions.

I attempted to develop methods that were independent of the Gaussian distribution but sensitive enough to correctly detect subtle differences between normal and diseased groups. In particular, I created or applied normal brain image atlases for detecting pathology in subjects across the lifecourse. While I have shown the potential of these atlases, they also had limitations.

10.6.3 Atlases

I created two atlases: one for subjects aged between 25 and 90 years and another for those aged 60-90 years. These will be useful for assessing brain structure in adult lifecourse or elderly adult studies (Buckner et al., 2004). However, there are large differences in brain structure between these ages (e.g. chapters 6, 7; Ge et al., 2002; Resnick et al., 2003; Farrell et al., 2009; Ferguson et al., 2010). This means that my atlases may not be as useful in studies with shorter age intervals. Further, my adult atlases were only in T1-wieghted sequences or images derived from T1 (GM images).

When we have collected sufficient data, I will create a third atlas from normal infant subjects. This will be used to highlight pathology in preterm infants. In the present work I used a mean-based normal atlas and so could determine the distance of preterm subjects from mean normal brain structure. I could not, however, determine whether these distances were out with the normal range. This, and validation of the other results reported here, will come with collection of more normal and pathological data from across the lifecourse.

10.6.4 Data and validation of statistical results

The data that I used in this work were cross-sectional. Previous work has shown that cross-sectional and longitudinal brain structure data produce similar results (Marcus et al., 2007c); but this may not always be the case (Freedman et al., 2007). The results in the present work may therefore not truly represent changes in brain structure but rather differences between ages and diagnoses. Nonetheless, I have provided methods that may be applied and validated in future longitudinal studies, e.g. the effect of percentile rank in conversion to AD.

I attempted to validate my results by repeating analyses between different samples. Although this is considered the most reliable way to make causal inferences in social and epidemiological science (Meehl, 1978; Cohen, 1994; Freedman, 2010), I only had two normal ageing samples (combined n=236) in which to validate my results. This is clearly not adequate to make reliable population inferences. There are large databanks of brain images from older subjects, e.g. the Rotterdam (Cees De Groot et al., 2000) and Framingham (Jeerakathil et al., 2004) study databanks, but these were not available to me during this work. In the future I may seek collaborations with these groups. Such collaborations are in line with the growing

recognition of the need for large, publicly accessible repositories of brain image data and statistical models (Poline and Poldrack, 2013).

The data available for many of the subjects I studied did not include risk factors such as blood pressure and number of prescriptions. Although high blood pressure and several prescriptions are common in older people and these data were collected in the original studies (Marcus et al., 2007c), these actual measures were not available to me. Without having these measures I was unable to determine potential sources of the variance in normal ageing brain structure. Further, I was not able to determine whether non-normal distributions were due to the inclusion of subjects with and without disease that effects brain structure, e.g., hypertension. It may be that a group of aged subjects who were entirely disease-free, i.e., "super-normal" (Mazziotta et al., 2009), have distributions that more closely follow the normal distribution. Furthermore, early neurodegenerative disease (that has not yet reached the point of abnormal clinical scores) may have contributed to the bimodal distributions I found e.g., Figure 58. Future work with longitudinal conversion rates may be used to exclude data from these subjects retrospectively. Subsequent repeated assessments of nonparametric versus parametric methods (in data where early neurodegenerative disease is retrospectively excluded) will determine whether it is an unknown mixture of populations that is causing normal ageing brain MRI data to appear non-normally distributed.

Future work with the BRAINS bank, and the considerable medical metadata it contains (chapter 9), may be able to determine whether unknown mixtures of populations are causing normal ageing brain MRI data to appear non-normally distributed. Other future work is described in the next section.

10.7 Future work

Limitations in the present work provide several directions for future work. The immediate priority is to implement parametric and nonparametric models of longitudinal whole brain volume changes (with the same subjects over time). This will determine if mean changes in brain volume are reflected in subjects at percentile ranks surrounding the mean, e.g. the 2.5^{th}, …, 97.5^{th} percentile ranks. Further, as data become available, I will test whether percentile rank level of brain volume change is

an early marker for the development of cognitive decline and neurodegenerative disease. Given that these diseases are often associated with subtle changes in brain structure (Shenton et al., 2001; Farrell et al., 2009; Ferguson et al., 2010), I will also test whether percentile ranks of regional and voxel tissue volumes can provide more sensitive markers.

The possibility for subtle differences between normal and diseased groups led me to not using conventional nonlinear registration or voxel smoothing. In future I will perform these common image processing steps and determine their effect on voxel distributions and statistical modelling results. Conversely, because I found it to be used most often in brain image analyses (Chapter 4), I only assessed the Gaussian distribution in this work. Future work may assess the applicability of other distributions, e.g. "Chi-square".

I found the distributions of normal ageing brain structure to be highly variable (chapters 6-8). Potential sources of this variation were not investigated in the present work. Future work will investigate whether it is due to, for example, blood pressure, years of education, or socioeconomic status. Regardless of its' source, this variation suggests that a percentile rank–based reference for brain volumes may be useful to quantitatively rank individuals or to determine if the distributions of brain volumes in control groups are skewed. The relatively small sample used here (n=316 normal subjects 25-90 years), indicates that this reference will likely require much more data than are publicly available at present (chapters 2 and 8). The BRAINS bank (chapter 9) may allow us to determine if ~70 subjects per year (~6,370 subjects across the lifecourse) are sufficient.

10.8 Conclusion

Parametric methods have allowed us to understand general changes that occur in human brain structure through life. Nonparametric methods will allow us to take this understanding further: to understand the distributions and limits of these changes. This will ultimately improve diagnoses, treatment, and outcome of the neurodegenerative diseases that await us.

References

Adelman, G., Smith, B.H. (1987). Encyclopedia of Neuroscience. Boston: Birkhäuser.

Admiraal-Behloul, F., van den Heuvel, D.M.J., Olofsen, H., van Osch, M.J.P., van der Grond, J., van Buchem, M.A., Reiber, J.H.C. (2005). Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. *Neuroimage*. 28, 607-617.

ADNI. (2011). ADNI Publications: http://www.adni-info.org/Scientists/ADNIScientistsHome/ADNIPublications.aspx, last accessed 26 June 2011.

Allen Institute for Brain Science. (2011). Allen Human Brain Atlas: http://human.brain-map.org/mri_viewers/data, last accessed 28 August 2013.

Allen, J.S., Bruss, J., Brown, C.K., Damasio, H. (2005). Normal neuroanatomical variation due to age: the major lobes and a parcellation of the temporal region. *Neurobiology of Aging*. 26, 1245-1260.

Alzheimer's Disease International. (2013). Dementia statistics: http://www.alz.co.uk/research/statistics, last accessed 05 Sep 2013.

Anderson, W., Makins, G.H. (1889). CRANIO-CEREBRAL TOPOGRAPHY. *The Lancet*. 134, 61-64.

Aribisala, B.S., Valdés Hernández, M.C., Royle, N.A., Morris, Z., Muñoz Maniega, S., Bastin, M.E., Deary, I.J., Wardlaw, J.M. (2012). Brain atrophy associations with white matter lesions in the ageing brain: the Lothian Birth Cohort 1936. *European Radiology*. 10.1007/s00330-00012-02677-x.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., Wouters, P. (2004). Promoting Access to Public Research Data for Scientific, Economic, and Social Development. *Data Science Journal*. 3, 135-152.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*. 38, 95-113.

Ashburner, J., Friston, K. (2000). Voxel-Based Morphometry–The Methods. *Neuroimage*. 11, 805-821.

Ashburner, J., Friston, K. (2007). Voxel-Based Morphometry. *Statistical Parametric Mapping*. 92 - 100.

Ashburner, J., Friston, K.J. (2009). Computing average shaped tissue probability templates. *Neuroimage*. 45, 333-341.

Aubert-Broche, B., Fonov, V., Ghassemi, R., Narayanan, S., Arnold, D.L., Banwell, B., Sled, J.G., Collins, D.L. (2011). Regional brain atrophy in children with multiple sclerosis. *Neuroimage*. 58, 409-415.

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*. 12, 26-41.

Barkhof, F., Fox, N.C., Bastos-Leite, A.J., Scheltens, P. (2011). Neuroimaging in Dementia. Berlin Heidelberg: Springer-Verlag.

Berrington de Gonzalez, A., Mahesh, M., Kim, K.-P., Bhargavan, M., Lewis, R., Mettler, F., Land, C. (2009). Projected cancer risks from computed tomographic scans performed in the United States in 2007. *Archives of Internal Medicine*. 169, 2071-2077.

Biomedical Image Analysis Group Imperial College London. (2010). Information eXtraction from Images (IXI) dataset: http://www.brain-development.org/, last accessed 31 May 2011.
Biomedical Informatics Research Network (BIRN). (2009). Morphometry BIRN Multi-site Multi-session Structural MRI Data: http://www.birncommunity.org/data-catalog/morphometry-birn-multi-site-multi-session-structural-mri-data/, last accessed 10 October 2011.
Biomedical Informatics Research Network (BIRN). (2011). Function BIRN Data Repository: http://fbirnbdr.nbirn.net:8080/BDR/index.jsp, last accessed 07 October 2011.
Blatter, D., Bigler, E., Gale, S., Johnson, S., Anderson, C., Burnett, B., Parker, N., Kurth, S., Horn, S. (1995). Quantitative volumetric analysis of brain MR: normative database spanning 5 decades of life. *American Journal of Neuroradiology*. 16, 241.
Blennow, K., de Leon, M.J., Zetterberg, H. (2006). Alzheimer's disease. *The Lancet*. 368, 387-403.
Bloch, F. (1946). Nuclear Induction. *Physical review*. 70, 460-474.
Bonilha, L., Rorden, C., Appenzeller, S., Carolina Coan, A., Cendes, F., Min Li, L. (2006). Gray matter atrophy associated with duration of temporal lobe epilepsy. *Neuroimage*. 32, 1070-1079.
Box, G.E.P., Cox, D.R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*. 26, 211-252.
Braak, H., Braak, E. (1997). Staging of Alzheimer-related cortical destruction. *International Psychogeriatrics*. 9, 257-261.
BRAINnet Foundation. (2009). BRAINnet Database: http://www.brainnet.net/what-data-are-available/mri-fmri-and-dti/, last accessed 31 May 2011.
Breteler, M., Van Swieten, J., Bots, M., Grobbee, D., Claus, J., Van Den Hout, J., Van Harskamp, F., Tanghe, H., De Jong, P., Van Gijn, J., Hofman, A. (1994). Cerebral white matter lesions, vascular risk factors, and cognitive function in a population-based study. *Neurology*. 44, 1246-1252.
Brett, M., Christoff, K., Cusack, R., Lancaster, J. (2001). Using the Talairach atlas with the MNI template. *Neuroimage*. 13, S85.
Brett, M., Johnsrude, I.S., Owen, A.M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*. 3, 243-249.
Broca, P., Brabrook, E.W. (1881). [Memoir of Paul Broca]. *The Journal of the Anthropological Institute of Great Britain and Ireland*. 10, 242-261.
Brodmann, K. (1994). The Principles of Comparative Localisation in the Cerebral Cortex Based on Cytoarchitectonics. London: Smith-Gordon.
Brody, J.A. (1985). Prospects for an ageing population. *Nature*. 315, 463-466.
Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*. 13, 400-404.
Buckner, R.L., Head, D., Parker, J., Fotenos, A.F., Marcus, D., Morris, J.C., Snyder, A.Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage*. 23, 724-738.

Bull, J. (1961). History of Neuroradiology—The Presidential Address, Delivered at the British Institute of Radiology on October 20, 1960. *British Journal of Radiology*. 34, 69-84.
Bull, J. (1970). The History of Neuroradiology. *Proceedings of the Royal Society of Medicine*. 63, 637-643.
Bullitt, E., Smith, J.K., Lin, W. (2010). Designed Database of MR Brain Images of Healthy Volunteers: http://www.insight-journal.org/midas/community/view/21, last accessed 31 May 2011.
Cees De Groot, J., De Leeuw, F.-E., Oudkerk, M., Van Gijn, J., Hofman, A., Jolles, J., Breteler, M.M.B. (2000). Cerebral white matter lesions and cognitive function: The Rotterdam scan study. *Annals of Neurology*. 47, 145-151.
Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*. 114, 494-509.
Cocosco, C.A., Kollokian, V., Remi, K.S.K., Pike, G.B., Evans, A.C. (1997). Brainweb: Online interface to a 3D MRI simulated brain database. *Neuroimage*. 5(4 Pt 2), s425.
Coe, R. (2002). It's the Effect Size, Stupid: What effect size is and why it is important. In: Annual Conference of the British Educational Research Association University of Exeter, England.
Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciencies. New York: Psychology Press.
Cohen, J. (1994). The earth is round (p<. 05). *American Psychologist*. 49, 997-1003.
Collins, D.L., Holmes, C., Peters, T.M., Evans, A.C. (1995). Automatic 3D model based neuroanatomical segmentation. *Human Brain Mapping*. 3, 190-208.
Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N., Holmes, C.J., Evans, A.C. (1998). Design and construction of a realistic digital brain phantom. *Medical Imaging, IEEE Transactions on*. 17, 463-468.
Collins, F., Prabhakar, A. (2013). BRAIN Initiative Challenges Researchers to Unlock Mysteries of Human Mind: http://www.whitehouse.gov/blog/2013/04/02/brain-initiative-challenges-researchers-unlock-mysteries-human-mind, last accessed 05 Sep 13.
Courchesne, E., Chisum, H.J., Townsend, J., Cowles, A., Covington, J., Egaas, B., Harwood, M., Hinds, S., Press, G.A. (2000). Normal Brain Development and Aging: Quantitative Analysis at in Vivo MR Imaging in Healthy Volunteers. *Radiology*. 216, 672-682.
Davenport, C. (2013). Scientists win 2 bln euros to fight brain disease, study graphene. Reuters.
Dawber, T.R., Meadors, G.F., Moore Jr, F.E. (1951). Epidemiological approaches to heart disease: the Framingham Study. *American Journal of Public Health*. 41, 279.
Deary, I., Gow, A., Taylor, M., Corley, J., Brett, C., Wilson, V., Campbell, H., Whalley, L., Visscher, P., Porteous, D., Starr, J. (2007). The Lothian Birth Cohort 1936: a study to examine influences on cognitive ageing from age 11 to age 70 and beyond. *BMC Geriatrics*. 7, 28.
Deary, I.J., Der, G., Ford, G. (2001). Reaction times and intelligence differences: A population-based cohort study. *Intelligence*. 29, 389-399.
DeCarli, C., Massaro, J., Harvey, D., Hald, J., Tullberg, M., Au, R., Beiser, A., D'Agostino, R., Wolf, P.A. (2005). Measures of brain morphology and infarction in the Framingham Heart Study: establishing what is normal. *Neurobiology of Aging*. 26, 491-510.

DeCarlo, L.T. (1997). On the Meaning and Use of Kurtosis. *Psychological Methods*. 2, 292–307.

den Heijer, T., Launer, L.J., Prins, N.D., van Dijk, E.J., Vermeer, S.E., Hofman, A., Koudstaal, P.J., Breteler, M.M.B. (2005). Association between blood pressure, white matter lesions, and atrophy of the medial temporal lobe. *Neurology*. 64, 263-267.

Department of Radiology University of Pennsylvania. (2008). Brain-image Database (BRAID): https://www.rad.upenn.edu/sbia/braid/braid_web/index.html, last accessed 31 May 2011.

Derbyshire, D. (2003). More stars 'than grains of sand'. The Telegraph, 23 Jul 2003.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 31, 968-980.

Dickie, D.A., Job, D.E., Poole, I., Ahearn, T.S., Staff, R.T., Murray, A.D., Wardlaw, J.M. (2012). Do brain image databanks support understanding of normal ageing brain structure? A systematic review. *European Radiology*. 22, 1385-1394.

Dickson, J., Drury, H., Van Essen, D.C. (2001). The surface management system (SuMS) database: a surface–based database to aid cortical surface reconstruction, visualization and analysis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*. 356, 1277-1292.

Doyle, A.C. (1892). The Adventure of the Copper Beeches. In: The Strand Magazine.

Duvernoy, H.M. (1999). The Human Brain. New York: Springer.

Edland, S., Xu, Y., Plevak, M., O'Brien, P., Tangalos, E., Petersen, R., Jack, C. (2002). Total intracranial volume: normative values and lack of association with Alzheimer's disease. *Neurology*. 59, 272-274.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage*. 25, 1325-1335.

Ellis, K.A., Bush, A.I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N.T., Lenzo, N., Martins, R.N., Maruff, P. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics*. 21, 672-687.

Ellis, K.A., Rowe, C.C., Villemagne, V.L., Martins, R.N., Masters, C.L., Salvado, O., Szoeke, C., Ames, D. (2010). Addressing population aging and Alzheimer's disease through the Australian Imaging Biomarkers and Lifestyle study: Collaboration with the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's and Dementia*. 6, 291-296.

Elveback, L.R., Guillier, C.L., Keating Jr, F.R. (1970). Health, Normality, and the Ghost of Gauss. *JAMA: The Journal of the American Medical Association*. 211, 69-75.

EQUATOR Network. (2011). About EQUATOR: http://www.equator-network.org/about-equator/, last accessed 31 May 2011.

European Space Agency. (2013). How many stars are there in the Universe: http://www.esa.int/Our_Activities/Space_Science/Herschel/How_many_stars_are_there_in_the_Universe, last accessed 07 Sep 2013.

Evans, A.C. (2006a). The NIH MRI Study of Normal Brain Development. *Neuroimage*. 30, 184-202.
Evans, A.C. (2006b). The NIH MRI Study of Normal Brain Development database: https://nihpd.crbs.ucsd.edu/nihpd/info/index.html, last accessed 31 May 2011.
Evans, A.C., Collins, D.L., Mills, S.R., Brown, E.D., Kelly, R.L., Peters, T.M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. In: IEEE Nuclear Science Symposium and Medical Imaging Conference, vol. 3, pp 1813-1817. (1993). San Francisco, CA , USA.
Evans, A.C., Janke, A.L., Collins, D.L., Baillet, S. (2012). Brain templates and atlases. *Neuroimage*. 62, 911-922.
Farrall, A.J. (2006). Magnetic resonance imaging. *Practical Neurology*. 6, 318-325.
Farrell, C., Chappell, F., Armitage, P.A., Keston, P., MacLullich, A., Shenkin, S., Wardlaw, J.M. (2009). Development and initial testing of normal reference MR images for the brain at ages 65–70 and 75–80 years. *European Radiology*. 19, 177-183.
Ferguson, K.J., Wardlaw, J.M., MacLullich, A.M.J. (2010). Quantitative and Qualitative Measures of Hippocampal Atrophy Are Not Correlated in Healthy Older Men. *Journal of Neuroimaging*. 20, 157-162.
Filipek, P.A., Richelme, C., Kennedy, D.N., Caviness, V.S. (1994). The Young Adult Human Brain: An MRI-based Morphometric Analysis. *Cerebral Cortex*. 4, 344-360.
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S. (2002). Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron*. 33, 341-355.
Fischl, B., Van Der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D. (2004). Automatically parcellating the human cerebral cortex. *Cerebral Cortex*. 14, 11.
FMRIB. (2008). Atlases included with FSL: http://www.fmrib.ox.ac.uk/fsl/data/atlas-descriptions.html, last accessed 25 March 2011.
Folstein, M.F., Folstein, S.E., McHugh, P.R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 12, 189-198.
Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D.L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*. 47, Supplement 1, S102.
Fotenos, A.F., Snyder, A., Girton, L., Morris, J., Buckner, R. (2005). Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology*. 64, 1032-1039.
Foulkes, M., Wolf, P., Price, T., Mohr, J., Hier, D. (1988). The Stroke Data Bank: design, methods, and baseline characteristics. *Stroke*. 19, 547.
Fox, N.C., Crum, W.R., Scahill, R.I., Stevens, J.M., Janssen, J.C., Rossor, M.N. (2001). Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *The Lancet*. 358, 201-205.
Fox, N.C., Schott, J.M. (2004). Imaging cerebral atrophy: normal ageing to Alzheimer's disease. *The Lancet*. 363, 392-394.
Fox, P.T., Lancaster, J.L. (2002). Mapping context and content: the BrainMap model. *Nature Reviews Neuroscience*. 3, 319-321.

Freedman, D. (2010). Statistical Models and Causal Inference: A Dialogue with the Social Sciences Cambridge: Cambridge University Press.
Freedman, D., Lane, D. (1983). A Nonstochastic Interpretation of Reported Significance Levels. *Journal of Business & Economic Statistics*. 1, 292-298.
Freedman, D., Pisani, R., Purves, R. (2007). Statistics. New York: WW Norton.
Fritz, C.O., Morris, P.E., Richler, J.J. (2012). Effect Size Estimates: Current Use, Calculations, and Interpretation. *Journal of Experimental Psychology: General*. 141, 2-18.
Ge, Y., Grossman, R.I., Babb, J.S., Rabin, M.L., Mannon, L.J., Kolson, D.L. (2002). Age-related total gray matter and white matter changes in normal adult brain. Part I: volumetric MR imaging analysis. *American Journal of Neuroradiology*. 23, 1327-1333.
Geroldi, C., Pihlajamäki, M., Laakso, M., DeCarli, C., Beltramello, A., Bianchetti, A., Soininen, H., Trabucchi, M., Frisoni, G.B. (1999). APOE-ε4 is associated with less frontal and more medial temporal lobe atrophy in AD. *Neurology*. 53, 1825-1849.
Giorgio, A., Santelli, L., Tomassini, V., Bosnell, R., Smith, S., De Stefano, N., Johansen-Berg, H. (2010). Age-related changes in grey and white matter structure throughout adulthood. *Neuroimage*. 51, 943-951.
Good, C.D., Johnsrude, I.S., Ashburner, J., Henson, R.N.A., Friston, K.J., Frackowiak, R.S.J. (2001). A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. *Neuroimage*. 14, 21-36.
Google Scholar. (2011). Search within articles citing Marcus: Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults: http://scholar.google.co.uk/scholar?cluster=14063688880671780453&hl=en&as_sdt=2005&sciodt=1,5, last accessed 02 September 2011.
Gountouna, V., Job, D., McIntosh, A., Moorhead, T., Lymer, G., Whalley, H., Hall, J., Waiter, G., Brennan, D., McGonigle, D. (2010). Functional magnetic resonance imaging (fMRI) reproducibility and variance components across visits and scanning sites with a finger tapping task. *Neuroimage*. 49, 552-560.
Grabner, G., Janke, A.L., Budge, M.M., Smith, D., Pruessner, J., Collins, D.L. (2006). Symmetric atlasing and model based segmentation: an application to the hippocampus in older adults. *Medical Image Computing and Computer-Assisted Intervention*. 4191, 58-66.
Grady, C.L., Springer, M.V., Hongwanishkul, D., McIntosh, A.R., Winocur, G. (2006). Age-related changes in brain activity across the adult lifespan. *Journal of Cognitive Neuroscience*. 18, 227-241.
Gur, R.C., Mozley, P.D., Resnick, S.M., Gottlieb, G.L., Kohn, M., Zimmerman, R., Herman, G., Atlas, S., Grossman, R., Berretta, D., Erwin, R., Gur, R.E. (1991). Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*. 88, 2845-2849.
Hall, J., Whalley, H.C., McKirdy, J.W., Romaniuk, L., McGonigle, D., McIntosh, A.M., Baig, B.J., Gountouna, V.-E., Job, D.E., Donaldson, D.I., Sprengelmeyer, R., Young, A.W., Johnstone, E.C., Lawrie, S.M. (2008). Overactivation of Fear Systems to Neutral Faces in Schizophrenia. *Biological Psychiatry*. 64, 70-73.
Hammers, A., Allom, R., Koepp, M., Free, S., Myers, R., Lemieux, L., Mitchell, T., Brooks, D., Duncan, J. (2003). Three-dimensional maximum probability atlas

of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping*. 19, 224-247.
Hawkes, R., Holland, G., Moore, W., Worthington, B. (1980). Nuclear Magnetic Resonance (NMR) Tomography of the Brain: A Preliminary Clinical Assessment with Demonstration of Pathology. *Journal of Computer Assisted Tomography*. 4, 577-586.
Head, D., Snyder, A.Z., Girton, L.E., Morris, J.C., Buckner, R.L. (2005). Frontal-hippocampal double dissociation between normal aging and Alzheimer's disease. *Cerebral Cortex*. 15, 732-739.
Hill, D.L.G., Hawkes, D., Williams, S. (2010). Information eXtraction from Images (IXI): Details of Grant: http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=GR/S21533/02, last accessed 31 May 2011.
Hinke, R.M., Hu, X., Stillman, A.E., Kim, S.-G., Merkle, H., Salmi, R., Ugurbil, K. (1993). Functional magnetic resonance imaging of Broca's area during internal speech. *Neuroreport*. 4, 675-678.
Hoeffner, E., Mukherji, S., Srinivasan, A., Quint, D. (2012). Neuroradiology Back to the Future: Brain Imaging. *American Journal of Neuroradiology*. 33, 5-11.
Hofstadter, R. (1994). Felix Bloch 1905-1983: A Biographical Memoir. Washington D.C.: National Academy of Sciences.
Hogg, R.V., Tanis, E.A. (2010). Probability and Statistical Inference. Upper Saddle River, New Jersey: Pearson.
Holland, D., McEvoy, L.K., Dale, A.M., the Alzheimer's Disease Neuroimaging, I. (2012). Unbiased comparison of sample size estimates from longitudinal structural measures in ADNI. *Human Brain Mapping*. 33, 2586-2602.
Holland, G., Hawkes, R., Moore, W. (1980). Nuclear Magnetic Resonance (NMR) Tomography of the Brain: Coronal and Sagittal Sections. *Journal of Computer Assisted Tomography*. 4, 429-433.
Hollander, M., Wolfe, D.A. (1973). Nonparametric Statistical Methods. New York: John Wiley & Sons, Inc.
Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W., Evans, A.C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*. 22, 324.
Hounsfield, G.N. (1973). Computerized transverse axial scanning (tomography): Part 1. Description of system. *British Journal of Radiology*. 46, 1016-1022.
Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D.S., Calabresi, P.A., Pekar, J.J., van Zijl, P., Mori, S. (2008). Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage*. 39, 336-347.
Insel, T.R., Volkow, N.D., Landis, S.C., Li, T.K., Battey, J.F., Sieving, P. (2004). Limits to growth: why neuroscience needs large-scale science. *Nature Neuroscience*. 7, 426-427.
Isherwood, I. (2005). Sir Godfrey Hounsfield. *Radiology*. 234, 975-976.
Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*. 27, 685-691.
Jack Jr, C.R., Petersen, R.C., Xu, Y.C., Waring, S.C., O'Brien, P.C., Tangalos, E.G., Smith, G.E., Ivnik, R.J., Kokmen, E. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*. 49, 786–794.

Jasmin, L., Zieve, D. (2012). Circle of Willis: http://www.nlm.nih.gov/medlineplus/ency/imagepages/18009.htm, last accessed 09 July.
Jeerakathil, T., Wolf, P.A., Beiser, A., Massaro, J., Seshadri, S., D'Agostino, R.B., DeCarli, C. (2004). Stroke Risk Profile Predicts White Matter Hyperintensity Volume: The Framingham Study. *Stroke*. 35, 1857-1861.
Jenkinson, M., Bannister, P., Brady, M., Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*. 17, 825-841.
Jenkinson, M., Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*. 5, 143-156.
Jernigan, T.L., Archibald, S.L., Fennema-Notestine, C., Gamst, A.C., Stout, J.C., Bonner, J., Hesselink, J.R. (2001). Effects of age on tissues and regions of the cerebrum and cerebellum. *Neurobiology of Aging*. 22, 581-594.
Job, D.E., Whalley, H.C., McConnell, S., Glabus, M., Johnstone, E.C., Lawrie, S.M. (2002). Structural Gray Matter Differences between First-Episode Schizophrenics and Normal Controls Using Voxel-Based Morphometry. *Neuroimage*. 17, 880-889.
Johansen-Berg, H., Behrens, T.E.J., Sillery, E., Ciccarelli, O., Thompson, A.J., Smith, S.M., Matthews, P.M. (2005). Functional–anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. *Cerebral Cortex*. 15, 31.
Johnson, K.A., Becker, J.A. (1999). The Whole Brain Atlas: http://www.med.harvard.edu/AANLIB/home.html, last accessed 31 May 2011.
Jones, R. (2012). Neurogenetics: What makes a human brain? *Nature Reviews Neuroscience*. 13, 655-655.
Jovicich, J., Czanner, S., Han, X., Salat, D., van der Kouwe, A., Quinn, B., Pacheco, J., Albert, M., Killiany, R., Blacker, D., Maguire, P., Rosas, D., Makris, N., Gollub, R., Dale, A., Dickerson, B.C., Fischl, B. (2009). MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: Reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. *Neuroimage*. 46, 177-192.
Koenker, R., Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*. 46, 33-50.
Koslow, S.H. (2000). Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience*. 3, 863-865.
Kruggel, F. (2006). MRI-based volumetry of head compartments: normative values of healthy adults. *Neuroimage*. 30, 1-11.
Laboratory of Neuro Imaging. (2011a). Alzheimer's Disease Neuroimaging Initiative (ADNI): http://adni.loni.ucla.edu/, last accessed 31 May 2011.
Laboratory of Neuro Imaging. (2011b). LONI Image Data Archive (IDA) - Data Access: https://ida.loni.ucla.edu/services/Menu/IdaData.jsp?page=DATA&subPage=AVAILABLE_DATA, last accessed 12 May 2011.
Lancaster, J., Rainey, L., Summerlin, J., Freitas, C., Fox, P., Evans, A., Toga, A., Mazziotta, J. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Human Brain Mapping*. 5, 238-242.

Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*. 10, 120-131.

Lauterbur, P.C. (1973). Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature*. 242, 190-191.

Lehmann, M., Douiri, A., Kim, L.G., Modat, M., Chan, D., Ourselin, S., Barnes, J., Fox, N.C. (2010). Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. *Neuroimage*. 49, 2264-2274.

Lemaitre, H., Crivello, F., Grassiot, B., Alperovitch, A., Tzourio, C., Mazoyer, B. (2005). Age-and sex-related effects on the neuroanatomy of healthy elderly. *Neuroimage*. 26, 900-911.

Letovsky, S.I., Whitehead, S., Paik, C.H., Miller, G.A., Gerber, J., Herskovits, E.H., Fulton, T.K., Bryan, R.N. (1998). A brain image database for structure/function analysis. *American Journal of Neuroradiology*. 19, 1869-1877.

Lezak, M.D., Howieson, D.B., Loring, D.W. (2004). Neuropsychological Asessment: Oxford University Press.

Li, W., van Tol, M.-J., Li, M., Miao, W., Jiao, Y., Heinze, H.-J., Bogerts, B., He, H., Walter, M. (2012). Regional specificity of sex effects on subcortical volumes across the lifespan in healthy aging. *Human Brain Mapping*. n/a-n/a.

Long, X., Zhang, L., Liao, W., Jiang, C., Qiu, B., the Alzheimer's Disease Neuroimaging, I. (2012). Distinct laterality alterations distinguish mild cognitive impairment and Alzheimer's disease from healthy aging: Statistical parametric mapping with high resolution MRI. *Human Brain Mapping*. n/a-n/a.

Luengo-Fernandez, R., Leal, J., Gray, A. (2010). Dementia 2010: The economic burden of dementia and associated research funding in the United Kingdom. University of Oxford for the Alzheimer's Research Trust.

MacLullich, A., Ferguson, K., Deary, I., Seckl, J., Starr, J., Wardlaw, J. (2002). Intracranial capacity and brain volumes are associated with cognition in healthy elderly men. *Neurology*. 59, 169-174.

Maldjian, J.A., Baer, A., Pearson, K., Wagner, B. (2010). PickAtlas: http://fmri.wfubmc.edu/software/PickAtlas, last accessed 28 August 2013.

Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*. 19, 1233-1239.

Malone, I.B., Cash, D., Ridgway, G.R., MacManus, D.G., Ourselin, S., Fox, N.C., Schott, J.M. (2013). MIRIAD—Public release of a multiple time point Alzheimer's MR imaging dataset. *Neuroimage*. 70, 33-36.

Manolio, T.A., Kronmal, R.A., Burke, G.L., Poirier, V., O'Leary, D.H., Gardin, J.M., Fried, L.P., Steinberg, E.P., Bryan, R.N. (1994). Magnetic resonance abnormalities and cardiovascular disease in older adults. The Cardiovascular Health Study. *Stroke*. 25, 318-327.

Mansfield, P., Maudsley, A.A. (1977). Medical imaging by NMR. *British Journal of Radiology*. 50, 188-194.

Mansfield, P., Pykett, I. (1978). Biological and Medical Imaging by NMR. *Journal of Magnetic Resonance*. 29, 355-373.

Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L. (2010). Open Access Series of Imaging Studies (OASIS): Longitudinal MRI Data in Nondemented and Demented Older Adults. *Journal of Cognitive Neuroscience*. 22, 2677-2684.

Marcus, D.S., Olsen, T.R., Ramaratnam, M., Buckner, R.L. (2007a). The extensible neuroimaging archive toolkit. *Neuroinformatics*. 5, 11-33.

Marcus, D.S., Olsen, T.R., Ramaratnam, M., Buckner, R.L. (2011). XNAT Central: http://central.xnat.org/, last accessed 31 May 2011.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L. (2007b). Open Access Series of Imaging Studies (OASIS): http://www.oasis-brains.org/, last accessed 31 May 2011.

Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L. (2007c). Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*. 19, 1498-1507.

Massey Jr, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*. 46, 68-78.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B. (2001). A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London B Biological Sciences*. 356, 1293-1322.

Mazziotta, J., Woods, R., Iacoboni, M., Sicotte, N., Yaden, K., Tran, M., Bean, C., Kaplan, J., Toga, A. (2009). The myth of the normal, average human brain−The ICBM experience:(1) Subject screening and eligibility. *Neuroimage*. 44, 914-922.

McConnell Brain Imaging Centre of the Montreal Neurological Institute. (2004). BrainWeb: Simulated MRI Volumes for Normal Brain Database: http://mouldy.bic.mni.mcgill.ca/brainweb/selection_normal.html, last accessed 31 May 2011.

McConnell Brain Imaging Centre of the Montreal Neurological Institute. (2010). Atlases: http://www.bic.mni.mcgill.ca/ServicesAtlases/HomePage, last accessed 02 February 2012.

McEvoy, L.K., Fennema-Notestine, C., Roddey, J.C., Hagler, D.J., Holland, D., Karow, D.S., Pung, C.J., Brewer, J.B., Dale, A.M. (2009). Alzheimer Disease: Quantitative Structural Neuroimaging for Detection and Prediction of Clinical and Structural Changes in Mild Cognitive Impairment. *Radiology*. 251, 195-205.

McGraw, K.O., Wong, S. (1992). A Common Language Effect Size Statistic. *Psychological Bulletin*. 111, 361-365.

McIntosh, A., Moorhead, T., Job, D., Lymer, G., Maniega, S.M., McKirdy, J., Sussmann, J., Baig, B., Bastin, M., Porteous, D., Evans, K.L., Johnstone, E.C., Lawrie, S.M., Hall, J. (2007). The effects of a neuregulin 1 variant on white matter density and integrity. *Molecular Psychiatry*. 13, 1054-1059.

McIntosh, A., Whalley, H., McKirdy, J., Hall, J., Sussmann, J., Shankar, P., Johnstone, E., Lawrie, S. (2008). Prefrontal Function and Activation in

Bipolar Disorder and Schizophrenia. *American Journal of Psychiatry*. 165, 378-384.
McRobbie, D.W., Moore, E.A., Graves, M.J., Prince, M.R. (2006). MRI from Picture to Proton. Cambridge: Cambridge University Press.
McRobbie, D.W., Moore, E.A., Graves, M.J., Prince, M.R. (2007). MRI From Picture to Proton. Cambridge: Cambridge University Press.
Meehl, P.E. (1978). Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology. *Journal of Consulting and Clinical Psychology*. 46, 806-834.
Mega, M., Dinov, I., Mazziotta, J., Manese, M., Thompson, P., Lindshield, C., Moussai, J., Tran, N., Olsen, K. (2005). Automated brain tissue assessment in the elderly and demented population: construction and validation of a sub-volume probabilistic brain atlas. *Neuroimage*. 26, 1009-1018.
Milham, M., Buckner, R.L., Castellanos, F.X., Margulies, D., Zang, Y., Mennes, M., Gutman, D., Bangaru, S., Craddock, C., LaConte, S., Mostofsky, S., Villringer, A. (2011). 1000 Functional Connectomes Project: http://fcon_1000.projects.nitrc.org/index.html, last accessed October 10 2011.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. (2009a). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine*. 6, e1000097.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G. (2009b). The PRISMA Statement: http://www.prisma-statement.org/statement.htm, last accessed 31 May 2011.
Moorhead, T., Gountouna, V., Job, D., McIntosh, A., Romaniuk, L., Lymer, G., Whalley, H., Waiter, G., Brennan, D., Ahearn, T. (2009). Prospective multi-centre Voxel Based Morphometry study employing scanner specific segmentations: Procedure development using CaliBrain structural MRI data. *BMC Medical Imaging*. 9, 8.
Mori, S., Wakana, S., Van Zijl, P.C.M., Nagae-Poetscher, L. (2005). MRI atlas of human white matter: Elsevier Amsterdam, The Netherlands.
Morris, J.C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*. 43, 2412-2414.
Mortamet, B., Zeng, D., Gerig, G., Prastawa, M., Bullitt, E. (2005). Effects of healthy aging measured by intracranial compartment volumes using a designed MR brain database. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*. 3749, 383-391.
Mueller, S.G., Stables, L., Du, A.T., Schuff, N., Truran, D., Cashdollar, N., Weiner, M.W. (2007). Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4 T. *Neurobiology of Aging*. 28, 719-726.
Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's and Dementia: The Journal of the Alzheimer's Association*. 1, 55-66.
Nelson, H.E., Willison, J. (1991). National Adult Reading Test (NART): Test Manual. Windsor: Nfer-Nelson.
Neuroinformatics Research Group. (2011). Brainscape database: http://nrg.wustl.edu/project/data-sharing/, last accessed 10 October 2011.
Neuropsychiatric Imaging Research Laboratory. (2009). NIRL Imaging Database: http://nirlarc.duhs.duke.edu/nirle/, last accessed 31 May 2011.

Nichols, T.E., Holmes, A.P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*. 15, 1-25.
Nordenskjöld, R., Malmberg, F., Larsson, E.-M., Simmons, A., Brooks, S.J., Lind, L., Ahlström, H., Johansson, L., Kullberg, J. (2013). Intracranial volume estimated with commonly used methods could introduce bias in studies including brain volume measurements. *Neuroimage*. 83, 355-360.
Nowinski, W.L., Belov, D. (2003). The Cerefy Neuroradiology Atlas: a Talairach-Tournoux atlas-based tool for analysis of neuroimages available over the internet. *Neuroimage*. 20, 50-57.
Nyúl, L.G., Udupa, J.K. (1999). On Standardizing the MR Image Intensity Scale. *Magnetic Resonance in Medicine*. 42, 1072–1081.
O'Connor, J.P.B. (2003). Thomas Willis and the background to Cerebri Anatome. *Journal of the Royal Society of Medicine*. 96, 139-143.
Ono, M., Kubik, S., Abernathey, C.D. (1990). Atlas of the cerebral sulci. New York: Thieme.
Pakkenberg, B., Pelvig, D., Marner, L., Bundgaard, M.J., Gundersen, H.J.G., Nyengaard, J.R., Regeur, L. (2003). Aging and the human neocortex. *Experimental Gerontology*. 38, 95-99.
Peelle, J.E., Cusack, R., Henson, R.N.A. (2012). Adjusting for global effects in voxel-based morphometry: Gray matter decline in normal aging. *Neuroimage*. 60, 1503-1516.
Pernet, C., Poline, J., Demonet, J., Rousselet, G. (2009). Brain classification reveals the right cerebellum as the best biomarker of dyslexia. *BMC Neuroscience*. 10, 67.
Petersen, R., Aisen, P., Beckett, L., Donohue, M., Gamst, A., Harvey, D., Jack, C., Jagust, W., Shaw, L., Toga, A. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI). *Neurology*. 74, 201.
Poline, J.-B., Breeze, J.L., Ghosh, S.S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., Helmer, K.G., Marcus, D.S., Poldrack, R.A., Schwartz, Y., Ashburner, J., Kennedy, D.N. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics*. 6,
Poline, J.B., Poldrack, R.A. (2013). Introduction to the special issue: Toward a new era of databasing and data sharing for neuroimaging. *Neuroimage*. 82, 645-646.
Pooley, R.A. (2005). Fundamental Physics of MR Imaging. *Radiographics*. 25, 1087-1099.
Prayer, D., Kasprian, G., Krampl, E., Ulm, B., Witzani, L., Prayer, L., Brugger, P.C. (2006). MRI of normal fetal brain development. *European Journal of Radiology*. 57, 199-216.
Raz, N., Ghisletta, P., Rodrigue, K.M., Kennedy, K.M., Lindenberger, U. (2010). Trajectories of brain aging in middle-aged and older adults: regional and individual differences. *Neuroimage*. 51, 501-511.
Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K.M., Williamson, A., Acker, J.D. (2004). Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: replicability of regional differences in volume. *Neurobiology of Aging*. 25, 377-396.
Raz, N., Lindenberger, U., Rodrigue, K.M., Kennedy, K.M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., Acker, J.D. (2005). Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers. *Cerebral Cortex*. 15, 1676-1689.

Research Imaging Institute UTHSCSA. (2010). BrainMap database: http://brainmap.org/index.html, last accessed 31 May 2011.
Resnick, S.M., Pham, D.L., Kraut, M.A., Zonderman, A.B., Davatzikos, C. (2003). Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *The Journal of Neuroscience*. 23, 3295-3301.
Rohlfing, T., Zahr, N., Sullivan, E., Pfefferbaum, A. (2010). The SRI24 multichannel atlas of normal adult human brain structure. *Human Brain Mapping*. 31, 798-819.
Rorden, C. (2010). MRIcron Index: http://www.mccauslandcenter.sc.edu/mricro/mricron/index.html, last accessed 01 Sept 2013.
Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., Karnath, H.-O. (2012). Age-specific CT and MRI templates for spatial normalization. *Neuroimage*. 61, 957-965.
Rorden, C., Bonilha, L., Nichols, T.E. (2007). Rank-order versus mean based statistics for neuroimaging. *Neuroimage*. 35, 1531-1537.
Ross, M.L. (2012). The Oil Curse: How Petroleum Wealth Shapes the Development of Nations. New Jersey: Princeton University Press.
Rowland, A., Burns, M., Hartkens, T., Hajnal, J.V., Rueckert, D., Hill, D.L.G. (2004). Information extraction from images (IXI): Image processing workflows using a grid enabled image database. In: Distributed Databases in Medical Image Computing - MICCAI Rennes, France.
Salthouse, T.A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychological Bulletin*. 137, 753-784.
Sato, K., Taki, Y., Fukuda, H., Kawashima, R. (2003a). Japanese Reference Brains: http://www.idac.tohoku.ac.jp/JHBP/, last accessed 31 May 2011.
Sato, K., Taki, Y., Fukuda, H., Kawashima, R. (2003b). Neuroanatomical database of normal Japanese brains. *Neural networks*. 16, 1301-1310.
Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N., Fox, N.C. (2003). A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Archives of Neurology*. 60, 989.
Scarpazza, C., Sartori, G., De Simone, M.S., Mechelli, A. (2013). When the single matters more than the group: Very high false positive rates in single case Voxel Based Morphometry. *Neuroimage*. 70, 175-188.
Schmansky, N. (2010). Cortical Parcellation: http://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation, last accessed 22 March 2011.
Schwartz, A. (2013). First volley in the brain race? *Annals of Neurology*. 73, A7-A7.
Scott, A., Wild, C. (1991). Transformations and R^2. *The American Statistician*. 45, 127-129.
Selkoe, D.J. (2001). Alzheimer's disease: genes, proteins, and therapy. *Physiological Reviews*. 81, 741-766.
Selkoe, D.J. (2013). The Therapeutics of Alzheimer's Disease: Where We Stand and Where We are Heading. *Annals of Neurology*. Accepted Article (Accepted, unedited articles published online and citable. The final edited and typeset version of record will appear in future), page numbers not yet assigned.
Shattuck, D., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K., Poldrack, R., Bilder, R., Toga, A. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage*. 39, 1064-1080.

Shattuck, D.W., Leahy, R.M. (2002). BrainSuite: an automated cortical surface identification tool. *Medical Image Analysis*. 6, 129-142.
Shenkin, S.D., Bastin, M.E., MacGillivray, T.J., Deary, I.J., Starr, J.M., Wardlaw, J.M. (2003). Childhood and current cognitive function in healthy 80-year-olds: a DT-MRI study. *Neuroreport*. 14, 345-349.
Shenton, M., Dickey, C., Frumin, M., McCarley, R. (2001). A review of MRI findings in schizophrenia. *Schizophrenia Research*. 49, 1-52.
Shenton, M., Kikinis, R., McCarley, W., Saiviroonporn, P., Hokama, H., Robatino, A., Metcalf, D., Wible, C., Portas, C., Iosifescu, D. (1995). Harvard brain atlas: a teaching and visualization tool. In: Biomedical Visualisation, pp 10-17 Atlanta, GA: IEEE Computer Society.
Siegel, S. (1957). Nonparametric statistics. *The American Statistician*. 11, 13-19.
Siegel, S., Castellan, J. (1988). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.
Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., K oszewska, I., Wahlund, L.O., Soininen, H., Lovestone, S., Evans, A., Spenger, C. (2011). The AddNeuroMed framework for multi centre MRI assessment of Alzheimer's disease: experience from the first 24 months. *International Journal of Geriatric Psychiatry*. 26, 75-82.
Smith, S.M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*. 17, 143-155.
Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A., De Stefano, N. (2002). Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis. *Neuroimage*. 17, 479-489.
Sorensen, A.G., Wu, O. (2010). International Stroke Database: http://www.strokedatabase.org/, last accessed 31 May 2011.
Sowell, E., Thompson, P., Leonard, C., Welcome, S., Kan, E., Toga, A. (2004). Longitudinal Mapping of Cortical Thickness and Brain Growth in Normal Children. *Journal of Neuroscience*. 24, 8223-8231.
Sowell, E.R., Peterson, B.S., Thompson, P.M., Welcome, S.E., Henkenius, A.L., Toga, A.W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience*. 6, 309-315.
Sundsten, J.W. (1994). Digital Anatomist: Interactive Brain Atlas: http://www9.biostr.washington.edu/cgi-bin/DA/PageMaster?atlas:Neuroanatomy+ffpathIndex:Splash^Page+2, last accessed 25 March 2011.
Talairach, J., Tournoux, P. (1988). Co-planar Sterotactic Atlas of the Human Brain: 3-dimensional Proportional System: An Approach to Cerebral Imaging. Stuttgart: Georg Thieme Verlag.
Tang, Y., Hojatkashani, C., Dinov, I., Sun, B., Fan, L., Lin, X., Qi, H., Hua, X., Liu, S., Toga, A. (2010). The construction of a Chinese MRI brain atlas: A morphometric comparison study between Chinese and Caucasian cohorts. *Neuroimage*. 51, 33-41.
Technical University of Denmark Informatics. (2009). Brede Database: http://neuro.imm.dtu.dk/services/jerne/brede/, last accessed 31 May 2011.
Telegraph Media Group Limited. (2013). The world's biggest companies. The Telegraph, 18 Apr 2013.
The AddNeuroMed Study. (2011). Data Access: http://www.innomed-addneuromed.com/index.cfm?PID=108, last accessed 07 September 2011.

The Cardiovascular Health Study. (2010). Data Distribution Policy: http://www.chs-nhlbi.org/CHS_DistribPolicy.htm, last accessed 31 May 2011.
The Center for Morphometric Analysis MGH HMS. (2002). Internet Brain Volume Database: http://www.cma.mgh.harvard.edu/ibvd/, last accessed 31 May 2011.
The IMAGEN Consortium. (2011). Imagen Europe - 6th Framework Project: http://www.imagen-europe.com/, last accessed 31 May 2011.
Thompson, P.M., Hayashi, K.M., de Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W. (2003). Dynamics of gray matter loss in Alzheimer's disease. *The Journal of Neuroscience*. 23, 994-1005.
Thompson, P.M., Hayashi, K.M., de Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M., Toga, A.W. (2004). Mapping hippocampal and ventricular change in Alzheimer disease. *Neuroimage*. 22, 1754-1766.
Thompson, P.M., Mega, M.S., Woods, R.P., Zoumalan, C.I., Lindshield, C.J., Blanton, R.E., Moussai, J., Holmes, C.J., Cummings, J.L., Toga, A.W. (2001). Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cerebral Cortex*. 11, 1-16.
Thompson, P.M., Woods, R.P., Mega, M.S., Toga, A.W. (2000). Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. *Human Brain Mapping*. 9, 81-92.
Toga, A.W. (2002). Neuroimage databases: the good, the bad and the ugly. *Nature Reviews Neuroscience*. 3, 302-309.
Toga, A.W. (2009). LONI Image Data Archive User Manual. *Laboratory Of Neuro Imaging, UCLA*.
Toga, A.W., Thompson, P.M., Mori, S., Amunts, K., Zilles, K. (2006). Towards multimodal atlases of the human brain. *Nature Reviews Neuroscience*. 7, 952-966.
Tondelli, M., Wilcock, G.K., Nichelli, P., De Jager, C.A., Jenkinson, M., Zamboni, G. (2012). Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. *Neurobiology of Aging*. 33, 825.e825-825.e836.
Townsend, P. (1979). Poverty in the United Kingdom: A Survey of Household Resources and Standards of Living. Berkeley and Los Angeles: University of California Press.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*. 15, 273-289.
van Dokkum, P.G., Conroy, C. (2010). A substantial population of low-mass stars in luminous elliptical galaxies. *Nature*. 468, 940-942.
Van Essen, D.C. (2005). A population-average, landmark-and surface-based (PALS) atlas of human cerebral cortex. *Neuroimage*. 28, 635-662.
Van Essen Lab. (2001). Surface Management System Database: http://sumsdb.wustl.edu:8081/sums/index.jsp, last accessed 31 May 2011.
Van Horn, J.D., Grethe, J.S., Kostelec, P., Woodward, J.B., Aslam, J.A., Rus, D., Rockmore, D., Gazzaniga, M.S. (2001). The Functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large–scale databasing of neuroimaging studies. *Philosophical Transactions of the Royal Society of London B Biological Sciences*. 356, 1323-1339.

Van Horn, J.D., Grethe, J.S., Kostelec, P., Woodward, J.B., Aslam, J.A., Rus, D., Rockmore, D., Gazzaniga, M.S. (2007). The fMRI Data Center: http://www.fmridc.org/f/fmridc/index.html, last accessed 31 May 2011.

Van Horn, J.D., Toga, A.W. (2009). Is it time to re-prioritize neuroimaging databases and digital repositories? *Neuroimage*. 47, 1720-1734.

Vargha, A., Delaney, H.D. (2000). A Critique and Improvement of the *CL* Common Language Effect Size Statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*. 25, 101-132.

Vaughan, T., DelaBarre, L., Snyder, C., Tian, J., Akgun, C., Shrivastava, D., Liu, W., Olson, C., Adriany, G., Strupp, J., Andersen, P., Gopinath, A., van de Moortele, P.-F., Garwood, M., Ugurbil, K. (2006). 9.4T human MRI: Preliminary results. *Magnetic Resonance in Medicine*. 56, 1274-1282.

Walhovd, K.B., Fjell, A.M., Reinvang, I., Lundervold, A., Dale, A.M., Eilertsen, D.E., Quinn, B.T., Salat, D., Makris, N., Fischl, B. (2005a). Effects of age on volumes of cortex, white matter and subcortical structures. *Neurobiology of Aging*. 26, 1261-1270.

Walhovd, K.B., Fjell, A.M., Reinvang, I., Lundervold, A., Dale, A.M., Quinn, B.T., Salat, D., Makris, N., Fischl, B. (2005b). Neuroanatomical aging: Universal but not uniform. *Neurobiology of Aging*. 26, 1279-1282.

Wardlaw, J.M., Bastin, M.E., Valdés Hernández, M.C., Muñoz Maniega, S., Royle, N.A., Morris, Z., Clayden, J.D., Sandeman, E.M., Eadie, E., Murray, C., Starr, J.M., Deary, I.J. (2011). Brain aging, cognition in youth and old age and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. *International Journal of Stroke*. 6, 547-559.

Wardlaw, J.M., Murray, V., Berge, E., Del Zoppo, G.J. (2009). Thrombolysis for acute ischaemic stroke. *Cochrane Database of Systematic Reviews*. 4,

Wechsler, D. (1997a). WAIS-III, Wechsler Adult Intelligence Scale: Administration and Scoring Manual: Psychological Corporation.

Wechsler, D. (1997b). WMS-III: Wechsler memory scale administration and scoring manual: Psychological Corporation.

Wisse, L.E., Biessels, G.J., Heringa, S.M., Kuijf, H.J., Koek, D., Luijten, P.R., Geerlings, M.I. (2014). Hippocampal subfield volumes at 7T in early Alzheimer's disease and normal aging. *Neurobiology of Aging*. 35, 2039–2045.

Woods, R.P. (2005). MultiTracer: http://www.loni.ucla.edu/Software/MultiTracer, last accessed 31 August 2013.

Yang, X., Beason-Held, L., Resnick, S.M., Landman, B.A. (2011). Biological parametric mapping with robust and non-parametric statistics. *Neuroimage*. 57, 423-430.

Zamboni, G., de Jager, C.A., Drazich, E., Douaud, G., Jenkinson, M., Smith, A.D., Tracey, I., Wilcock, G.K. (2013). Structural and functional bases of visuospatial associative memory in older adults. *Neurobiology of Aging*. 34, 961-972.

Zhang, Y., Brady, M., Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*. 20, 45-57.

Zhao, L., Boucher, M., Rosa-Neto, P., Evans, A.C. (2012). Impact of scale space search on age- and gender-related changes in MRI-based cortical morphometry. *Human Brain Mapping*. doi: 10.1002/hbm.22050.

Ziegler, G., Dahnke, R., Jäncke, L., Yotter, R.A., May, A., Gaser, C. (2011). Brain structural trajectories over the adult lifespan. *Human Brain Mapping*. 33, 2377–2389.

Zigmond, A.S., Snaith, R.P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*. 67, 361-370.

Zilles, K., Amunts, K. (2010). Centenary of Brodmann's map — conception and fate. *Nature Reviews Neuroscience*. 11, 139-145.

Zimmerman, D.W. (1998). Invalidation of Parametric and Nonparametric Statistical Tests by Concurrent Violation of Two Assumptions. *The Journal of Experimental Education*. 67, 55-68.

APPENDIX I: Supplementary material

AI.1 Individual descriptions of each databank included in the systematic review

AI.1.1 ADNI databank

The ADNI was setup to determine which combination of neuroimaging, cerebral spinal fluid (CSF), and blood biomarkers provides the earliest and most accurate diagnosis and expected course of Alzheimer's disease (AD). Housed at the Laboratory of Neuroimaging (LONI) Image Data Archive (IDA), the ADNI databank contained serial MR brain images, separated by 6–12 months over 2–3 years, from approximately 229 normal, 398 mild cognitive impairment (MCI), and 192 AD subjects aged 55–90 years (normal subjects were aged 70–90 years; Table 3). Normal subjects may have had some medical problems common in ageing. The criteria for normality also included results from a battery of cognitive tests including the American National Adult Reading Test (ANART), Clinical Dementia Rating (CDR) and Mini Mental State Examination (MMSE). These results were included in the databank and subject retrieval parameters. Image analysis results, e.g. brain masks, were in the databank but statistical results were not. The LONI IDA also supported subject retrieval by clinico–demographic and scan acquisition parameters.

AI.1 2 AIBL databank

The AIBL study was designed to understand the pathology and early clinical manifestation of AD, improve the diagnosis of AD, and identify diet and lifestyle factors that are significant in the development of AD. Also housed at the LONI IDA, the AIBL databank contained serial MR brain images acquired from 177 normal, 57 MCI, and 53 AD subjects aged 60–100 years. Although medical and cognitive test results were part of the criteria for normality, the normal subjects were not representative of the normal ageing population because they were preferentially selected as APOE ε4 allele carriers (Ellis et al., 2010). The LONI IDA, which housed this databank, was further described previously (section 2.3.2.1).

AI.1 3 Designed Database of MR Brain Images of Healthy Volunteers

The Designed Database of MR Brain Images of Healthy Volunteers was created to assess the effects of healthy ageing on brain structure and provide references for the assessment of disease. It contained MR brain images from 100 normal subjects aged 18–72 years (15 subjects were aged ≥60 years; Table 3). These subjects had "no history of diabetes, hypertension, head trauma, psychiatric disease, or other symptoms or history likely to affect the brain". Therefore, they may not have been representative of the entire normal ageing population. I did not find cognitive test results to be in the criteria for normality. Subject demographics age, gender, race and handedness were in the database and these could be used to retrieve subjects.

AI.1.4 The fMRIDC

The fMRIDC was created to allow the functional magnetic resonance imaging (fMRI) research community to validate methods and hypotheses and perform meta–analyses of a large number of peer reviewed studies. It stored structural MR brain images as well as the fMRI from these studies. I did not have access to the overall composition of the databank, e.g. I could not determine the total number of subjects, but found 66 normal subjects aged ≥60 years. These subjects had no history of stroke, heart attack or psychiatric disorder but had medical characteristics common in ageing e.g. hypertension, arthritis. Fifty of these subjects were also in the OASIS and XNAT Central databanks. Subjects could be retrieved by study title, author, keywords, abstract or "special collections" (selected novel datasets) but not directly by subject clinico–demographic, imaging or cognitive parameters.

AI.1.5 IXI dataset

The IXI dataset was acquired to develop computer aided diagnostics of MR brain images. It contained MR brain images from 593 normal, healthy subjects aged 19–86 years (197 subjects were aged ≥60 years; Table 3). I did not find the criteria for normality and medical and cognitive test results were not in the dataset. Subject demographics such as age, gender, weight, ethnicity and qualification were in the dataset and these could be used to retrieve subjects.

AI.1.6 ICBM databank

The ICBM study will develop a probabilistic atlas and reference system for the normal human brain throughout the lifespan (Mazziotta et al., 2001). The ICBM databank, also housed at the LONI IDA, contained MR brain images from 851 normal subjects aged 18–90 years (approximately 76 subjects were aged ≥60 years; Table 3). The criteria for normality included results from several medical and cognitive tests such as the MMSE. As the criteria for normality were the same regardless of age, the older subjects may not have been representative of the entire normal ageing population e.g. subjects with any prescription medications (with some exceptions such as antibiotics) or hypertension were excluded regardless of age. The LONI IDA, which housed this databank, was further described previously (section 2.3.2.1).

AI.1.7 OASIS: Cross–sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults

The OASIS cross–sectional databank was created to provide the data needed for widespread study of ageing and dementia and to develop new MR brain image analysis techniques. It contained MR brain images from 316 normal and 100 demented subjects aged 18–96 years (98 normal subjects were aged ≥60 years; Table 3). Some of the normal older subjects had hypertension and treated diabetes. The criteria for normality also included MMSE score and CDR. The databank supported subject retrieval by clinico–demographics, cognitive test results (CDR and MMSE), scanning acquisition parameters, and parameters derived from analyses of brain images, e.g. brain tissue volume normalised by head size. Although these image analysis results were included in the databank, statistical results were not.

AI.1.8 OASIS: Longitudinal MRI Data in Nondemented and Demented Older Adults

The OASIS longitudinal databank was created for reasons similar to the OASIS cross–sectional databank. It contained serial MR brain images, acquired over two or more sessions and separated by at least 1 year, from 86 normal (14 of which later converted to dementia) and 64 demented subjects aged 60–96 years. Many of the older subjects in the OASIS cross–sectional databank were also in this longitudinal databank but were assigned new subject identifiers. The criteria for normality and

subject retrieval parameters were the same as in the OASIS cross–sectional databank (section 2.3.2.7).

AI.1.9 The Extensible Neuroimaging Archive Toolkit (XNAT) Central databank

The XNAT Central databank was designed to allow secure and quality controlled data (medical image and metadata) sharing among local colleagues, external collaborators and the broader neuroscience community. It contained over 3000 subjects from approximately 200 medical imaging studies, including the OASIS data. In addition to the OASIS subjects, I found 9 subjects (with brain images) aged ≥60 years. However, these subjects were not normal (they were neurosurgery patients). I did not have access to the remaining subjects or criteria for normality. Subjects could be retrieved by the parameters in the OASIS cross–sectional databank among others, e.g. regional brain volumes.

AI.2 Criteria for normality in each normal ageing brain volume statistical model

AI.2.1 Allen, et al. (2005)

All subjects were right–handed (and had no left–handedness in first degree relatives), healthy, and had no history of neurological or psychiatric illness. Subjects over 60 years were interviewed for general health status and medication usage. Included subjects had no clinical history of heart disease, hypertension, diabetes, or any other common age–related disease.

AI.2.2 Blatter et al. (1995)

There were no neurologic or neuropsychiatric examinations but all subjects answered a questionnaire and were excluded if they had a history of: previous head injury causing loss of consciousness; any disease affecting the nervous system, including dementia or psychiatric illness; or alcohol or drug abuse. Based on an assessment of motor function, 95% of subjects were considered left hemisphere dominant.

AI.2.3 Courchesne, et al. (2000)

Based on their responses to questionnaires, included subjects showed no evidence of developmental, educational, medical, psychological, or psychiatric abnormalities or deficiencies. Subjects older than 50 years were tested with the California Verbal Learning Test and had normal or above scores.

AI.2.4 DeCarli, et al. (2005)

By a review of medical history and physical examination, subjects with stroke, dementia, multiple sclerosis, or other clinically evident neurological conditions were excluded. Included subjects had various diseases of the vascular system such as hypertension and peripheral vascular disease.

AI.2.5 Fotenos, et al. (2005)

Older subjects with severe comorbidities such as depression or disabling stroke were excluded.

AI.2.6 Ge, et al. (2002)

A structured clinical interview excluded subjects with present or past neuropsychiatric illnesses, abuse of alcohol or illicit drugs, or psychiatric disease in first–degree relatives. Included subjects showed normal findings in a physical examination and none reported a history of any serious medical condition.

AI.2.7 Giorgio, et al. (2010)

Included subjects had no history of psychiatric or neurological disease or substance abuse. They did not show white matter lesions or overt abnormalities on MR brain images, e.g. infarct, vascular malformation, or tumour.

AI.2.8 Good, et al. (2001)

Subjects with neurological, medical, psychiatric conditions, or migraine were excluded. Included subjects had no history of: alcohol intake of more than 30 units per week or more than 10 units within 48h of scan; head trauma requiring medical attention; cognitive difficulties; treated hypertension (leading to a small number of subjects aged > 70 years). Measures were obtained by a questionnaire but no cognitive tests were performed.

AI.2.9 Gur, et al. (1991)

Subjects with central nervous system infection, seizure, head trauma with loss of consciousness, cerebrovascular disease, cognitive decline, alcohol or other substance abuse, psychiatric disorders, or psychiatric disorders in first–degree relatives were excluded.

AI.2.10 Jernigan, et al. (2001)

Included subjects were screened for current and significant medical, psychiatric, intellectual, or neurological disorders. Middle–aged and elderly subjects had a physical examination, showed normal cognitive function and were living independently. Elderly subjects with medical conditions common with age, e.g. hypertension and cardiac conditions, were not excluded if stable and well controlled.

AI.2.11 Kruggel (2006)

Included subjects passed a brief history and physical inspection by a physician.

AI.2.12 Li et al. (2012)

Subjects with below normal cognitive scores, e.g. MMSE < 28, hypertension, prescription medications, or a history of central nervous system (CNS) diseases were excluded.

AI.2.13 Long et al. (2012)

Medical history was reviewed and subjects with endocrinal, neurological, or psychiatric illnesses were excluded.

AI.2.14 Peelle et al. (2012)

Self reports and a standard pre–MRI screening questionnaire were used to assess neurological difficulty and general health. Included subjects had no neurological difficulty and were in good general health at the time of scan.

AI.2.15 Sowell, et al. (2003)

Subjects with a history of concussion, substance abuse, or seizure disorder were excluded. Included subjects showed no neurological impairments, psychiatric illness,

history of learning disability, or development delay according to a structured diagnostic interview.

AI.2.16 Walhovd, et al. (2005)

Included subjects were right–handed, felt well and healthy, had normal or normal–corrected sight, did not use a hearing aid, and did not have an illness known to affect the central nervous system e.g. hypothyroidism, multiple sclerosis, Parkinson's disease, stroke, or head injury. Excluded subjects had either health or cognitive problems according to a structured interview, Beck depression inventory (BDI) score >14, mini–mental state examination MMSE) score <26, and >1 SD below the mean Wechsler abbreviated intelligence scale (WASI) score.

AI.2.17 Ziegler et al. (2011)

All subjects were reported to be "normal and healthy" but no criteria for this were found; 79.9% of subjects had advanced or even higher educational levels.

AI.3 Differences between parametric and nonparametric distributions of brain volumes across early to middle adult life

The differences between parametric and nonparametric distributions of brain volumes that were illustrated in the figures in Chapter 7, "The accuracy of Gaussian distributions of regional brain volumes across early to middle adult life" are listed in the following tables.

Table AI.3.1. Differences in parametric and nonparametric distributions of whole brain volumes between age groups

Volume	Comparison	P2.5TH	NP2.5TH	% Err	P16TH	NP16TH	% Err	P50TH	NP50TH	% Err	P84TH	NP84TH	% Err	P97.5TH	NP97.5TH	% Err
CSF	"1–2"	–3.9E–03	–1.1E–02	172	–3.9E–03	–3.0E–03	25	–3.9E–03	1.5E–03	–139	–3.9E–03	–7.4E–03	88	–3.9E–03	–4.6E–03	18
	"1–3"	–4.2E–03	–1.1E–02	152	–4.2E–03	–4.7E–03	11	–4.2E–03	–5.4E–04	87	–4.2E–03	–5.1E–03	21	–4.2E–03	–1.2E–04	97
	"1–4"	–1.3E–02	–3.1E–02	145	–1.3E–02	–1.6E–02	28	–1.3E–02	–9.2E–03	27	–1.3E–02	–9.9E–03	22	–1.3E–02	2.3E–03	–118
	"2–3"	–3.1E–04	0.0E+00	Inf	–3.1E–04	–1.7E–03	458	–3.1E–04	–2.1E–03	564	–3.1E–04	2.3E–03	–825	–3.1E–04	4.5E–03	–1543
	"2–4"	–8.8E–03	–2.0E–02	133	–8.8E–03	–1.3E–02	51	–8.8E–03	–1.1E–02	23	–8.8E–03	–2.5E–03	71	–8.8E–03	6.9E–03	–178
	"3–4"	–8.5E–03	–2.0E–02	142	–8.5E–03	–1.2E–02	36	–8.5E–03	–8.7E–03	3	–8.5E–03	–4.8E–03	43	–8.5E–03	2.4E–03	–128
GM	"1–2"	8.2E–03	–2.9E–03	–135	8.2E–03	9.3E–03	14	8.2E–03	1.3E–02	59	8.2E–03	1.2E–02	41	8.2E–03	1.5E–02	87
	"1–3"	1.1E–02	1.7E–02	54	1.1E–02	1.8E–02	54	1.1E–02	7.7E–03	32	1.1E–02	1.5E–02	32	1.1E–02	1.8E–02	63
	"1–4"	1.5E–02	3.9E–03	74	1.5E–02	1.6E–02	3	1.5E–02	1.6E–02	5	1.5E–02	2.0E–02	33	1.5E–02	3.2E–02	110
	"2–3"	3.2E–03	2.0E–02	546	3.2E–03	8.2E–03	160	3.2E–03	–5.4E–03	–270	3.2E–03	3.3E–03	6	3.2E–03	3.1E–03	0
	"2–4"	6.9E–03	6.8E–03	1	6.9E–03	6.3E–03	9	6.9E–03	2.8E–03	59	6.9E–03	8.4E–03	22	6.9E–03	1.6E–02	138
	"3–4"	3.7E–03	–1.4E–02	–463	3.7E–03	–1.9E–03	–152	3.7E–03	8.2E–03	119	3.7E–03	5.1E–03	36	3.7E–03	1.3E–02	255
WM	"1–2"	–4.3E–03	–6.5E–04	85	–4.3E–03	1.0E–04	–102	–4.3E–03	–4.6E–03	8	–4.3E–03	–8.1E–03	89	–4.3E–03	–4.8E–03	13
	"1–3"	–7.1E–03	–1.4E–03	80	–7.1E–03	2.7E–03	–138	–7.1E–03	–4.8E–03	32	–7.1E–03	–1.9E–02	160	–7.1E–03	–2.9E–02	307
	"1–4"	–2.4E–03	3.2E–03	–232	–2.4E–03	8.7E–04	–136	–2.4E–03	5.7E–04	–124	–2.4E–03	–1.1E–02	374	–2.4E–03	–8.7E–03	262
	"2–3"	–2.8E–03	–7.4E–04	74	–2.8E–03	2.6E–03	–191	–2.8E–03	–1.7E–04	94	–2.8E–03	–1.0E–02	267	–2.8E–03	–2.4E–02	750
	"2–4"	1.9E–03	3.8E–03	103	1.9E–03	7.7E–04	59	1.9E–03	5.2E–03	177	1.9E–03	–3.3E–03	–275	1.9E–03	–3.9E–03	–305
	"3–4"	4.7E–03	4.6E–03	4	4.7E–03	–1.8E–03	–139	4.7E–03	5.4E–03	14	4.7E–03	7.2E–03	52	4.7E–03	2.0E–02	330

Note: P=parametric/ NP=nonparametric percentile rank; GM=grey matter; WM=white matter; CSF=cerebrospinal fluid; %Err=percent error; Comp=comparison; 1=26–35 years; 2=36–45 years; 3=46–55 years; 4=56–65 years; Inf=infinite (divided by zero).

Table AI.3.2. Differences in parametric and nonparametric distributions of left hippocampal complex volumes between age groups

Volume	Comp	P2.5TH	NP2.5TH	% Err	P16TH	NP16TH	% Err	P50TH	NP50TH	% Err	P84TH	NP84TH	% Err	P97.5TH	NP97.5TH	% Err
Left Amyg	"1–2"	1.9E–06	–1.4E–05	–825	1.9E–06	3.6E–06	92	1.9E–06	–7.3E–07	–139	1.9E–06	–3.1E–06	–263	1.9E–06	1.3E–05	570
	"1–3"	–7.6E–06	–5.1E–05	570	–7.6E–06	–1.1E–05	50	–7.6E–06	–6.2E–06	19	–7.6E–06	3.5E–06	–147	–7.6E–06	–9.5E–06	25
	"1–4"	–1.8E–05	–7.0E–05	295	–1.8E–05	1.7E–06	–110	–1.8E–05	–1.8E–05	4	–1.8E–05	–8.2E–06	54	–1.8E–05	–2.6E–05	47
	"2–3"	–9.5E–06	–3.7E–05	291	–9.5E–06	–1.5E–05	58	–9.5E–06	–5.4E–06	43	–9.5E–06	6.6E–06	–170	–9.5E–06	–2.2E–05	134
	"2–4"	–2.0E–05	–5.7E–05	187	–2.0E–05	–1.9E–06	90	–2.0E–05	–1.8E–05	10	–2.0E–05	–5.1E–06	74	–2.0E–05	–3.9E–05	98
	"3–4"	–1.0E–05	–1.9E–05	91	–1.0E–05	1.3E–05	–228	–1.0E–05	–1.2E–05	20	–1.0E–05	–1.2E–05	15	–1.0E–05	–1.7E–05	64
Left Hippo	"1–2"	–4.7E–05	7.2E–05	–254	–4.7E–05	–5.9E–05	25	–4.7E–05	–9.0E–05	92	–4.7E–05	–5.2E–05	10	–4.7E–05	2.2E–05	–147
	"1–3"	–7.3E–05	–1.3E–04	77	–7.3E–05	–1.0E–04	38	–7.3E–05	–9.3E–05	27	–7.3E–05	–5.4E–05	26	–7.3E–05	3.9E–05	–153
	"1–4"	–8.1E–05	–7.6E–05	6	–8.1E–05	–2.2E–05	73	–8.1E–05	–1.2E–04	46	–8.1E–05	–1.3E–04	59	–8.1E–05	3.5E–05	–143
	"2–3"	–2.6E–05	–2.0E–04	675	–2.6E–05	–4.2E–05	61	–2.6E–05	–2.4E–06	91	–2.6E–05	–2.5E–06	90	–2.6E–05	1.7E–05	–164
	"2–4"	–3.4E–05	–1.5E–04	338	–3.4E–05	3.7E–05	–209	–3.4E–05	–2.8E–05	19	–3.4E–05	–7.7E–05	127	–3.4E–05	1.3E–05	–138
	"3–4"	–8.0E–06	5.3E–05	–762	–8.0E–06	7.9E–05	–1087	–8.0E–06	–2.5E–05	216	–8.0E–06	–7.4E–05	833	–8.0E–06	–4.0E–06	50
Left Phippo	"1–2"	–1.3E–05	–2.0E–06	84	–1.3E–05	4.9E–05	–484	–1.3E–05	–2.5E–05	95	–1.3E–05	–6.8E–05	432	–1.3E–05	–4.0E–05	211
	"1–3"	–1.3E–05	5.4E–05	–516	–1.3E–05	–4.2E–05	228	–1.3E–05	–3.2E–05	146	–1.3E–05	1.7E–05	–229	–1.3E–05	6.9E–05	–632
	"1–4"	–1.9E–05	–5.5E–06	71	–1.9E–05	5.0E–05	–366	–1.9E–05	–1.5E–05	18	–1.9E–05	–5.4E–05	190	–1.9E–05	–3.6E–05	93
	"2–3"	–1.0E–07	5.6E–05	–55804	–1.0E–07	–9.2E–05	91446	–1.0E–07	–6.7E–06	6600	–1.0E–07	8.5E–05	–84860	–1.0E–07	1.1E–04	–108539
	"2–4"	–5.9E–06	–3.5E–06	41	–5.9E–06	6.3E–07	–111	–5.9E–06	9.7E–06	–264	–5.9E–06	1.4E–05	–335	–5.9E–06	3.7E–06	–162
	"3–4"	–5.8E–06	–5.9E–05	920	–5.8E–06	9.2E–05	–1689	–5.8E–06	1.6E–05	–382	–5.8E–06	–7.1E–05	1122	–5.8E–06	–1.0E–04	1706

Note: P=parametric/ NP=nonparametric percentile rank; %Err=percent error; Amyg=amygdala; Hippo=hippocampus; Phippo Gy=parahippocampal gyrus; Comp=comparison; 1=26–35 years; 2=36–45 years; 3=46–55 years; 4=56–65 years.

Table AI.3.3. Differences in parametric and nonparametric distributions of right hippocampal complex volumes between age groups

Volume	Comp	P2.5TH	NP2.5TH	% Err	P16TH	NP16TH	% Err	P50TH	NP50TH	% Err	P84TH	NP84TH	% Err	P97.5TH	NP97.5TH	% Err
Right Amyg	"1–2"	–2.2E–05	–3.8E–06	83	–2.2E–05	–1.5E–05	32	–2.2E–05	–2.6E–05	19	–2.2E–05	–1.0E–05	53	–2.2E–05	–1.6E–05	26
	"1–3"	–3.1E–05	–7.7E–05	153	–3.1E–05	–2.7E–05	11	–3.1E–05	–2.3E–05	26	–3.1E–05	–2.7E–05	12	–3.1E–05	–3.3E–05	9
	"1–4"	–2.0E–05	–5.3E–05	168	–2.0E–05	–8.3E–06	58	–2.0E–05	–9.2E–06	53	–2.0E–05	–1.8E–05	10	–2.0E–05	–4.9E–05	148
	"2–3"	–8.9E–06	–7.3E–05	727	–8.9E–06	–1.3E–05	41	–8.9E–06	3.3E–06	–137	–8.9E–06	–1.7E–05	86	–8.9E–06	–1.7E–05	95
	"2–4"	1.9E–06	–4.9E–05	–2660	1.9E–06	6.5E–06	237	1.9E–06	1.7E–05	764	1.9E–06	–7.5E–06	–492	1.9E–06	–3.3E–05	–1819
	"3–4"	1.1E–05	2.4E–05	126	1.1E–05	1.9E–05	76	1.1E–05	1.3E–05	23	1.1E–05	9.0E–06	17	1.1E–05	–1.6E–05	–245
Right Hippo	"1–2"	–5.6E–05	–5.2E–05	7	–5.6E–05	–7.7E–05	37	–5.6E–05	–6.1E–05	10	–5.6E–05	–6.7E–05	21	–5.6E–05	–5.0E–06	91
	"1–3"	–6.2E–05	–7.8E–05	26	–6.2E–05	–1.3E–04	109	–6.2E–05	–5.7E–05	7	–6.2E–05	–3.9E–05	37	–6.2E–05	4.6E–06	–107
	"1–4"	–5.8E–05	–5.1E–05	13	–5.8E–05	–1.2E–04	111	–5.8E–05	–1.2E–05	80	–5.8E–05	–1.3E–04	116	–5.8E–05	–1.7E–06	97
	"2–3"	–5.9E–06	–2.6E–05	340	–5.9E–06	–5.2E–05	788	–5.9E–06	4.2E–06	–171	–5.9E–06	2.9E–05	–584	–5.9E–06	9.6E–06	–263
	"2–4"	–2.4E–06	1.2E–06	–151	–2.4E–06	–4.7E–05	1840	–2.4E–06	5.0E–05	–2166	–2.4E–06	–5.9E–05	2348	–2.4E–06	3.3E–06	–238
	"3–4"	3.5E–06	2.7E–05	677	3.5E–06	5.8E–06	67	3.5E–06	4.5E–05	1196	3.5E–06	–8.7E–05	–2595	3.5E–06	–6.3E–06	–280
Right Phippo	"1–2"	–3.2E–05	–2.8E–05	15	–3.2E–05	–6.7E–05	106	–3.2E–05	5.0E–08	–100	–3.2E–05	–5.3E–05	63	–3.2E–05	–9.0E–05	178
	"1–3"	–3.4E–05	–3.6E–05	4	–3.4E–05	–8.0E–05	134	–3.4E–05	–7.2E–06	79	–3.4E–05	–6.3E–05	84	–3.4E–05	–1.0E–05	69
	"1–4"	–4.1E–05	–1.9E–05	53	–4.1E–05	–1.5E–05	63	–4.1E–05	–1.8E–05	56	–4.1E–05	–3.3E–05	18	–4.1E–05	–8.1E–05	98
	"2–3"	–1.8E–06	–8.1E–06	352	–1.8E–06	–1.3E–05	639	–1.8E–06	–7.3E–06	303	–1.8E–06	–1.0E–05	458	–1.8E–06	7.9E–05	–4514
	"2–4"	–8.4E–06	8.3E–06	–199	–8.4E–06	5.2E–05	–716	–8.4E–06	–1.8E–05	114	–8.4E–06	1.9E–05	–332	–8.4E–06	9.0E–06	–207
	"3–4"	–6.6E–06	1.6E–05	–349	–6.6E–06	6.5E–05	–1086	–6.6E–06	–1.1E–05	62	–6.6E–06	3.0E–05	–547	–6.6E–06	–7.0E–05	967

Note: P=parametric/ NP=nonparametric percentile rank; %Err=percent error; Amyg=amygdala; Hippo=hippocampus; Phippo Gy=parahippocampal gyrus; Comp=comparison; 1=26–35 years; 2=36–45 years; 3=46–55 years; 4=56–65 years.

Table AI.3.4. Differences in parametric and nonparametric distributions of left thalamus and basal ganglia volumes between age groups

Volume	Comp	P2.5TH	NP2.5TH	% Err	P16TH	NP16TH	% Err	P50TH	NP50TH	% Err	P84TH	NP84TH	% Err	P97.5TH	NP97.5TH	% Err
Left Caudate	"1–2"	9.0E–07	–5.6E–06	–727	9.0E–07	–1.2E–05	–1404	9.0E–07	9.8E–06	990	9.0E–07	–1.4E–05	–1697	9.0E–07	–2.1E–05	–2442
	"1–3"	3.1E–05	3.1E–06	90	3.1E–05	1.5E–05	51	3.1E–05	3.1E–05	0	3.1E–05	1.1E–05	65	3.1E–05	6.5E–05	106
	"1–4"	5.3E–05	–6.0E–05	–212	5.3E–05	9.1E–06	83	5.3E–05	8.0E–05	50	5.3E–05	–4.2E–06	–108	5.3E–05	1.6E–04	195
	"2–3"	3.0E–05	8.8E–06	71	3.0E–05	2.7E–05	11	3.0E–05	2.1E–05	29	3.0E–05	2.5E–05	17	3.0E–05	8.6E–05	182
	"2–4"	5.3E–05	–5.4E–05	–203	5.3E–05	2.1E–05	60	5.3E–05	7.0E–05	34	5.3E–05	1.0E–05	81	5.3E–05	1.8E–04	240
	"3–4"	2.2E–05	–6.3E–05	–385	2.2E–05	–6.1E–06	–128	2.2E–05	4.9E–05	122	2.2E–05	–1.5E–05	–168	2.2E–05	9.3E–05	321
Left Putamen	"1–2"	4.0E–05	–4.0E–05	–199	4.0E–05	8.7E–06	78	4.0E–05	5.6E–05	40	4.0E–05	4.3E–05	8	4.0E–05	1.4E–04	238
	"1–3"	1.1E–04	5.7E–05	46	1.1E–04	1.4E–04	33	1.1E–04	1.2E–04	12	1.1E–04	9.2E–05	13	1.1E–04	1.3E–04	21
	"1–4"	6.9E–05	–1.7E–05	–124	6.9E–05	1.5E–04	114	6.9E–05	6.8E–05	2	6.9E–05	9.3E–06	87	6.9E–05	6.2E–05	10
	"2–3"	6.5E–05	9.6E–05	47	6.5E–05	1.3E–04	101	6.5E–05	6.2E–05	5	6.5E–05	4.9E–05	26	6.5E–05	–7.9E–06	–112
	"2–4"	2.9E–05	2.3E–05	20	2.9E–05	1.4E–04	379	2.9E–05	1.2E–05	60	2.9E–05	–3.4E–05	–217	2.9E–05	–7.3E–05	–352
	"3–4"	–3.6E–05	–7.3E–05	101	–3.6E–05	7.6E–06	–121	–3.6E–05	–5.0E–05	38	–3.6E–05	–8.3E–05	127	–3.6E–05	–6.5E–05	79
Left Thalamus	"1–2"	–2.3E–06	3.0E–04	–13147	–2.3E–06	1.3E–04	–5540	–2.3E–06	–3.5E–05	1416	–2.3E–06	–1.7E–04	7207	–2.3E–06	–1.5E–05	554
	"1–3"	–8.5E–05	6.1E–05	–172	–8.5E–05	–3.5E–05	59	–8.5E–05	–1.1E–04	23	–8.5E–05	–1.7E–04	101	–8.5E–05	–4.0E–05	54
	"1–4"	–2.1E–04	–1.2E–04	39	–2.1E–04	–7.3E–05	65	–2.1E–04	–1.7E–04	17	–2.1E–04	–3.0E–04	48	–2.1E–04	–4.8E–04	135
	"2–3"	–8.3E–05	–2.4E–04	187	–8.3E–05	–1.6E–04	92	–8.3E–05	–7.0E–05	15	–8.3E–05	–3.6E–06	96	–8.3E–05	–2.5E–05	70
	"2–4"	–2.0E–04	–4.2E–04	109	–2.0E–04	–2.0E–04	3	–2.0E–04	–1.4E–04	33	–2.0E–04	–1.4E–04	33	–2.0E–04	–4.7E–04	131
	"3–4"	–1.2E–04	–1.9E–04	55	–1.2E–04	–3.8E–05	68	–1.2E–04	–6.6E–05	45	–1.2E–04	–1.3E–04	10	–1.2E–04	–4.4E–04	270

Note: P=parametric/ NP=nonparametric percentile rank; %Err=percent error; Comp=comparison; 1=26–35 years; 2=36–45 years; 3=46–55 years; 4=56–65 years.

Table AI.3.1.5. Differences in parametric and nonparametric distributions of right thalamus and basal ganglia volumes between ages

Volume	Comp	P2.5TH	NP2.5TH	% Err	P16TH	NP16TH	% Err	P50TH	NP50TH	% Err	P84TH	NP84TH	% Err	P97.5TH	NP97.5TH	% Err
Right Caudate	"1–2"	–3.4E–05	–9.4E–05	175	–3.4E–05	–6.6E–05	92	–3.4E–05	–4.1E–05	21	–3.4E–05	–3.8E–05	10	–3.4E–05	1.2E–04	–448
	"1–3"	–5.8E–05	–6.1E–05	5	–5.8E–05	–3.8E–05	34	–5.8E–05	–6.4E–05	11	–5.8E–05	–7.7E–05	33	–5.8E–05	–2.1E–05	64
	"1–4"	–1.0E–05	–2.5E–04	2283	–1.0E–05	–6.0E–05	475	–1.0E–05	6.6E–05	–730	–1.0E–05	–4.3E–06	59	–1.1E–05	1.1E–04	–1169
	"2–3"	–2.4E–05	3.3E–05	–240	–2.4E–05	2.8E–05	–217	–2.4E–05	–2.3E–05	4	–2.4E–05	–3.9E–05	66	–2.4E–05	–1.4E–04	493
	"2–4"	2.4E–05	–1.6E–04	–759	2.4E–05	5.3E–06	78	2.4E–05	1.1E–04	353	2.4E–05	3.3E–05	41	2.4E–05	–6.8E–06	–128
	"3–4"	4.7E–05	–1.9E–04	–500	4.7E–05	–2.2E–05	–147	4.7E–05	1.3E–04	175	4.7E–05	7.3E–05	53	4.7E–05	1.3E–04	182
Right Putamen	"1–2"	4.9E–05	–1.9E–05	–139	4.9E–05	–4.9E–06	–110	4.9E–05	5.9E–05	21	4.9E–05	9.1E–06	81	4.9E–05	1.0E–04	105
	"1–3"	1.6E–04	9.2E–05	41	1.6E–04	1.3E–04	15	1.6E–04	1.9E–04	20	1.6E–04	1.4E–04	11	1.6E–04	2.2E–04	38
	"1–4"	1.0E–04	–3.0E–05	–130	1.0E–04	3.8E–05	63	1.0E–04	1.5E–04	44	1.0E–04	8.3E–05	18	1.0E–04	2.2E–04	114
	"2–3"	1.1E–04	1.1E–04	3	1.1E–04	1.4E–04	28	1.1E–04	1.3E–04	19	1.1E–04	1.3E–04	20	1.1E–04	1.2E–04	7
	"2–4"	5.3E–05	–1.1E–05	–121	5.3E–05	4.3E–05	19	5.3E–05	8.7E–05	64	5.3E–05	7.4E–05	40	5.3E–05	1.2E–04	121
	"3–4"	–5.4E–05	–1.2E–04	125	–5.4E–05	–9.4E–05	73	–5.4E–05	–4.1E–05	25	–5.4E–05	–5.5E–05	1	–5.4E–05	2.0E–06	–104
Right Thalamus	"1–2"	–3.5E–05	1.2E–04	–430	–3.5E–05	6.5E–05	–282	–3.5E–05	–6.6E–05	85	–3.5E–05	–7.5E–05	111	–3.5E–05	–2.0E–04	455
	"1–3"	–1.2E–04	9.2E–05	–178	–1.2E–04	–7.6E–05	35	–1.2E–04	–1.3E–04	9	–1.2E–04	–1.4E–04	20	–1.2E–04	–3.7E–04	214
	"1–4"	–2.2E–04	–2.2E–04	2	–2.2E–04	–1.7E–04	23	–2.2E–04	–2.4E–04	8	–2.2E–04	–2.4E–04	6	–2.2E–04	–1.7E–04	23
	"2–3"	–8.2E–05	–2.5E–05	70	–8.2E–05	–1.4E–04	71	–8.2E–05	–6.3E–05	24	–8.2E–05	–6.7E–05	19	–8.2E–05	–1.7E–04	110
	"2–4"	–1.9E–04	–3.4E–04	78	–1.9E–04	–2.4E–04	26	–1.9E–04	–1.8E–04	7	–1.9E–04	–1.6E–04	14	–1.9E–04	2.4E–05	–113
	"3–4"	–1.1E–04	–3.1E–04	193	–1.1E–04	–9.6E–05	9	–1.1E–04	–1.1E–04	7	–1.1E–04	–9.5E–05	11	–1.1E–04	2.0E–04	–286

Note: P=parametric/ NP=nonparametric percentile rank; %Err=percent error; Comp=comparison; 1=26–35 years; 2=36–45 years; 3=46–55 years; 4=56–65 years.

APPENDIX II: Technical material

AII.1 BRAINS schema

The BRAINS schema was developed mainly by Dr Dominic Job, Dr Susan Shenkin, Dr David Rodriguez Gonzalez, and Prof Joanna Wardlaw, The University of Edinburgh.

Entity /Attributes	Description	Status
Subject		
BraINS Trial /Study ID	Reference to Trial /Study	invisible
BraINS Subject ID	Alphanumeric identifier	foreground
Sex	Male; Female; Other	foreground
Demographics		
BraINS Subject ID	Reference to Subject	invisible
Age at data collection	Age in years at physical data collection time	foreground
Age at scan time	Age in years at image data collection time	foreground
Age at cognitive data collection time	Age in years at cognitive data collection time	foreground
Handedness	Left; Right; Both	foreground
First language	Language code	foreground
Age at leaving school	Age in years	foreground
Years of formal education	Number of years	foreground
Highest qualification	Qualification code	foreground
Occupation	Occupation code	foreground
Ethnicity	Ethnicity code	foreground
Country of Birth	Country of Birth	foreground
Country of Data Collection	Country of Data Collection	foreground
Demographics Notes	Text	background
Measurements		
BraINS Subject ID	Reference to Subject	invisible
Age at data collection	Age in years	foreground
Blood pressure - systolic	Reading in mmHg	foreground
Blood pressure - systolic	Reading in mmHg	background
Blood pressure - diastolic	Reading in mmHg	foreground
Blood pressure - diastolic	Reading in mmHg	background
Pulse	Reading in BPM	background
Weight	Weight in kg	foreground
Height	Height in cm	foreground
BMI	Score	foreground
Medical examination normal	Y/N	REMOVE
Measurements Notes	Text	background
Medical History		
BraINS Subject ID	Reference to Subject	invisible
Age at data collection	Age in years	foreground
Method of Collection	Self Report; Medical Records; Interview	background

Prior head trauma	Y/N	background
Prior ischaemic stroke	Y/N	foreground
Prior TIA	Y/N	foreground
Prior Stroke or TIA	Y/N	foreground
Prior haemorrhagic stroke	Y/N	foreground
Peripheral vascular disease	Y/N	foreground
Peripheral vascular disease duration	Duration in years	background
IHD /MI	Y/N	foreground
IHD /MI duration	Duration in years	background
AF clinical history	Y/N	foreground
AF on ECG	Y/N	foreground
AF duration	Duration in years	background
Hypercholesterolaemia	Y/N	foreground
Hypercholesterolaemia duration	Duration in years	background
Hypertension	Y/N	foreground
Hypertension duration	Duration in years	background
Diabetes Type I	Y/N	foreground
Diabetes Type I duration	Duration in years	background
Diabetes Type II	Y/N	foreground
Diabetes Type II duration	Duration in years	background
Epilepsy	Y/N	foreground
Epilepsy duration	Duration in years	background
Dementia	Y/N	foreground
Dementia duration	Duration in years	background
Cognitive Impairment	Y/N	foreground
Cognitive Impairment duration	Duration in years	background
Depression	Y/N	foreground
Depression duration	Duration in years	background
Schizophrenia	Y/N	foreground
Schizophrenia duration	Duration in years	background
Parkinson's	Y/N	foreground
Parkinson's duration	Duration in years	background
Peripheral arthritis/osteoarthritis	Y/N	foreground
Peripheral arthritis/osteoarthritis	Duration in years	background
Medical History Notes	Text	background

Medication summary

BraINS Subject ID	Reference to Subject	invisible
Any anti-platelet	Y/N	foreground
Any statin	Y/N	foreground
Any anti-hypertensive	Y/N	foreground
Any psychiatric drug	Y/N	foreground
Any dementia drug	Y/N	foreground

Medications

BraINS Subject ID	Reference to Subject	invisible
Age at data collection	Age in years	background
Aspirin	Y/N	background
Warfarin	Y/N	background
Low dose heparin	Y/N	REMOVE
High dose heparin	Y/N	REMOVE
Anticoagulant (unspecified)	Y/N	background

Clopidogrel	Y/N	background
Dipyridamole	Y/N	background
Antiplatelet (unspecified)	Y/N	background
Beta blocker	Y/N	background
ACE inibitor or angiotensin II receptor antagonist	Y/N	background
Calcium antagonist	Y/N	background
Diuretic	Y/N	background
Statin	Y/N	background
Cardiovascular drug (unspecified)	Y/N	background
Benzodiazepine	Y/N	background
Antidepressant (unspecified)	Y/N	background
Antipsychotic (unspecified)	Y/N	background
Memantine	Y/N	background
Cholinesterase inhibitor	Y/N	background
Antidementia (unspecified)	Y/N	background
Thyroxine	Y/N	background
Analgesic (excluding NSAID)	Y/N	background
NSAID	Y/N	background
Steroid (excluding cream/inhaler)	Y/N	background
Anti-inflammatory (unspecified)	Y/N	background
Oral hypoglycaemic	Y/N	background
Insulin	Y/N	background
Inhaler (steroid)	Y/N	background
Inhaler (cholinergic; beta agonist; other)	Y/N	background
Proton pump inhibitor /H2 blocker	Y/N	background
Parkinson's medication (unspecified)	Y/N	background
Anticholinergic	Y/N	background
Total number oral medications	Number of meds	foreground
Medications Notes	Text	background

Risk Factors		
BraINS Subject ID	Reference to Subject	invisible
Age at data collection	Age in years	foreground
Smoking	Never; Previous; Current	foreground
Smoking - age started	Age in years	background
Smoking - age stopped	Age in years	background
Smoking -	Number of cigarettes	background
Alcohol - units /week	Number of units	foreground
Family history of stroke	Y/N	background
Family history of peripheral vascular disease	Y/N	background
Family history of IHD /MI	Y/N	background
Family history of hypercholesterolaemia	Y/N	background
Family history of hypertension	Y/N	background
Family history of diabetes Type I	Y/N	background
Family history of diabetes Type II	Y/N	background
Family history of epilepsy	Y/N	background
Family history of dementia	Y/N	background
Family history of cognitive impairment	Y/N	background
Family history of depression	Y/N	background
Family history of schizophrenia	Y/N	background
Family history of Parkinson's	Y/N	background
Diet history available	Y/N	background

Risk Factors Notes	Text	background

Clinical data - Scan Findings

BraINS Subject ID	Reference to Subject	invisible
Age at scan	Age in years	foreground
Focal lesions - Infarct - Cortical	Y/N	background
Focal lesions - Infarct - Striatocapsular	Y/N	background
Focal lesions - Infarct - Borderzone	Y/N	background
Focal lesions - Infarct - Lacunar	Y/N	background
Focal lesions - Infarct - Brainstem /Cerebellum	Y/N	background
Focal lesions - Haemorrhage	Y/N	background
Focal lesions - Non-Stroke	Y/N	background
Cerebral tumour	Y/N	foreground
EPVS - basal ganglia	Score (0-4)	background
EPVS - centrum semiovale	Score (0-4)	background
EPVS - hippocampus	Score (0-4)	background
BMBs	Score	foreground
Atrophy score	Score	foreground
Atrophy ventricle	Score	foreground
Atrophy sulci	Score	foreground
White matter lesions - Fazekas (PVH)	Score (0-3)	foreground
White matter lesions - Fazekas (DWMH)	Score (0-3)	foreground
White matter lesions - Longstreth (PVH+DWMH)	Score (0-8)	foreground
White matter lesions - Wahlund (PVH+DWMH)	Score (1-3)	background
Brain volume	Volume in mm3	foreground
ICV	Volume in mm3	background
CSF vol	Volume in mm3	background
Brain vol as % of ICV	Volume in mm3	background
WML vol	Volume in mm3	background
WML vol as % of ICV	Volume in mm3	background
Intracranial area	Area in mm2	foreground
Incidental findings	Text	background
Incidental findings	Y/N	foreground
Scan Findings Notes	Text	background
Neuroradiology report	Text	foreground
Normal Scan		Y/N

Clinical data - Cognitive Tests

BraINS Subject ID	Reference to Subject	invisible
Age at test	Age in years at time of cognitive test	foreground
MMSE	Score	foreground
Moray House Test	Score	background
NART	Score	foreground
Ravens Progressive Matrices	Score	background
Wechsler Matrices	Score	background
Wechsler (WASI)	Score	background
Wechsler (WRAT)	Score	background
Digit symbol	Score	background
Auditory verbal learning test	Score	background
IDED	Score	background
CVLT	Score	background
Extended Rivermead memory test	Score	background

HADS - anxiety	Score	foreground
HADS - depression	Score	foreground
MoCA	Score	foreground
ACE-R	Score	foreground
Verbal Fluency	Score	background
Logical Memory	Score	background
Cognitive Tests Notes	Text	background
Clinical data - Functional Tests		
BraINS Subject ID	Reference to Subject	invisible
Townsend	Score	background
Barthel	Score	background
Grip strength	Score	background
6 meter walk	Score	background
Clinical data - Laboratory Tests		
BraINS Subject ID	Reference to Subject	invisible
Hb	Score	background
HbA1c	Score	background
cholesterol	Score	background
triglyc	Score	background
creatinine	Score	background
Albumin	Score	background
CRP	Score	background

Note: The third column ("Status") refers to the visibility of a variable in the databank.

APPENDIX III: Key outputs and awards

AIII.1 List of key publications and presentations

DICKIE, D.A., et al.:

(Accepted). Variance in brain volume with advancing age: implications for defining the limits of normality. *PLOS ONE*.

(Submitted). Evidence of non-normal distributions in brain imaging data from normal subjects: implications for diagnosis of disease. *ISMRM 2014 Annual Meeting*.

(2013). How normal is this brain? Development and testing of a new MR template for voxel–based brain ranking. *Proceedings of the 19th Annual Meeting of the Organization for Human Brain Mapping*. Seattle, USA.

(2013). Development of automatic, voxel–based brain ranking. *Anne Rowling Regenerative Neurology Clinic launch meeting*. Edinburgh, UK.

(2013). Brain Images of Normal Subjects (BRAINS) Bank. *NHS Lothian Annual Research Conference*. Edinburgh, UK. **Poster Competition First Prize.**

(2012). Distinguishing normal and pathological ageing brain structures with data not statistics. *CCACE 5th Annual Research Day*. Edinburgh, UK. **Young Scientist Poster Award.**

(2012). A databank, rather than statistical, model of normal ageing brain structure to indicate pathology. *Proceedings of the 18th Annual Meeting of the Organization for Human Brain Mapping*. Beijing, China.

(2012). Do brain image databanks support understanding of normal ageing brain structure? A systematic review. *Eur Radiol*. 22 (7), 1385–1394. **Received editorial:** Barkhof, F. (2012). Making better use of our brain MRI research data.

Valdés Hernández, M. C., …, DICKIE, D.A., et al.:

(2013). Differentiation of Calcified Regions and Iron Deposits in the Ageing Brain on Conventional Structural MR Images. *J. Magn. Reson. Imaging*.

(2012). Close correlation between quantitative and qualitative assessments of white matter lesions. *Neuroepidemiology*. 40 (1), 13–22.

(2012). Comparison of ageing–related white matter lesion quantification by volume and visual rating scores. *American Stroke Association International Stroke Conference*. New Orleans, Louisiana, USA.

AIII.2 List of funding awards

Anne Rowling Regenerative Neurology Clinic, April 2013, 250 GBP.

Postdoctoral and Early Career Researcher Exchange Fund, Scottish Funding Council, July 2012, 5,000 GBP.

Travel grant, Guarantors of Brain, April 2012, 800 GBP.

AIII.3 Copies of first author publications

Eur Radiol
DOI 10.1007/s00330-012-2392-7

COMPUTER APPLICATIONS

Do brain image databanks support understanding of normal ageing brain structure? A systematic review

David Alexander Dickie · Dominic E. Job · Ian Poole · Trevor S. Ahearn · Roger T. Staff · Alison D. Murray · Joanna M. Wardlaw

Received: 25 October 2011 / Revised: 5 December 2011 / Accepted: 29 December 2011

Abstract
Objective To document accessible magnetic resonance (MR) brain images, metadata and statistical results from normal older subjects that may be used to improve diagnoses of dementia.
Methods We systematically reviewed published brain image databanks (print literature and Internet) concerned with normal ageing brain structure.
Results From nine eligible databanks, there appeared to be 944 normal subjects aged ≥60 years. However, many subjects were in more than one databank and not all were fully representative of normal ageing clinical characteristics. Therefore, there were approximately 343 subjects aged ≥60 years with metadata representative of normal ageing, but only 98 subjects were openly accessible. No databank had the range of MR image sequences, e.g. T2*, fluid-attenuated inversion recovery (FLAIR), required to effectively characterise the features of brain ageing. No databank supported random subject retrieval; therefore, manual selection bias and errors may occur in studies that use these subjects as controls. Finally, no databank stored results from statistical analyses of its brain image and metadata that may be validated with analyses of further data.
Conclusion Brain image databanks require open access, more subjects, metadata, MR image sequences, searchability and statistical results to improve understanding of normal ageing brain structure and diagnoses of dementia.
Key Points
- *We reviewed databanks with structural MR brain images of normal older people.*
- *Among these nine databanks, 98 normal subjects ≥60 years were openly accessible.*
- *None had all the required sequences, random subject retrieval or statistical results.*
- *More access, subjects, sequences, metadata, searchability and results are needed.*
- *These may improve understanding of normal brain ageing and diagnoses of dementia.*

Keywords Magnetic resonance imaging · Normality · Databanks · Review · Brain disease

D. A. Dickie · D. E. Job · J. M. Wardlaw
Division of Clinical Neurosciences, Western General Hospital,
Brain Research Imaging Centre (BRIC), University of Edinburgh,
Crewe Road,
Edinburgh EH4 2XU, UK

T. S. Ahearn · R. T. Staff · A. D. Murray
Aberdeen Biomedical Imaging Centre, University of Aberdeen,
Lilian Sutton Building, Foresterhill,
Aberdeen AB25 2ZD, UK

D. A. Dickie (✉) · D. E. Job · T. S. Ahearn · R. T. Staff · A. D. Murray · J. M. Wardlaw
Scottish Imaging Network,
A Platform for Scientific Excellence (SINAPSE) collaboration,
1 George Square,
Edinburgh EH8 9JZ, UK
e-mail: d.a.dickie@sms.ed.ac.uk

I. Poole
Toshiba Medical Visualisation Systems Europe, Ltd.,
Bonnington Bond, 2 Anderson Place,
Edinburgh EH6 5NP, UK

Introduction

Normal ageing and dementia are associated with brain tissue loss (atrophy) measured by magnetic resonance (MR) imaging [1–7]. The progressive economic and human burden of dementia, and the likely limited effect of future treatments beyond the early clinical stages, necessitates earlier and

Published online: 22 February 2012

more accurate diagnoses of pathological atrophy [8–10]. Better understanding of the effects of normal ageing on brain tissue, in visual or computational assessment tools, may afford earlier and more accurate diagnoses of dementia and other age-related neurological disorders [1].

The variation of brain tissue loss in normal subjects increases with advancing age [1, 4, 5]. Thus, one-off studies including relatively small numbers of subjects may not be reliable [11–13]. Indeed, these studies are in disagreement as to whether atrophy (of grey matter) is slowed in old age [6], constant across adulthood [3], or increased in old age [2]. Reports of correlations of normal brain volumes and cognitive measures are also inconsistent [14].

With large volumes of data, brain image databanks may facilitate a better understanding of the variation in structure of the normal ageing brain [11–13, 15–19], e.g. results from statistical analyses (such as correlation coefficients of brain volumes and age) can be stored and validated with analyses of further data. Image data should include a range of MR sequences, such as T1, T2, T2* and fluid-attenuated inversion recovery (FLAIR), to effectively describe the features of normal brain ageing [20]. These databanks should support random and stratified random, i.e. random within study-specific constraints such as brain structure volumes and clinico-demographics, image (subject) retrieval so to remove manual selection bias and provide appropriately matched controls in studies of ageing.

Studies of ageing should also address the issue of what is "normal" [21]. Many older people without neurological disorder have clinical characteristics that may affect brain structure, such as hypertension, diabetes, arthritis and medication use that may not be regarded as "normal" in younger populations [1, 21–27]. Therefore, these clinical characteristics should be considered part of the "effects of normal ageing" and subjects with them, if otherwise cognitively normal, should probably not be excluded from normal ageing brain image studies and databanks. Instead, it should be possible to say in which way subjects are normal, i.e. whether they are ageing "successfully" (without these clinical characteristics) or "usually" (with these characteristics) [28], and hence their inclusion and categorisation in normal ageing brain image studies and databanks should be supported by thorough cognitive and medical test results (metadata) [21]. This is particularly true if these subjects are to be used as controls in studies of dementia and related disorders where cognitive state greatly influences diagnoses [29, 30].

It is apparent that brain image databanks have the potential to support studies of age-related neurological disorders and better understanding of normal ageing brain structure [19]. A systematic review may determine if they have yet realised this potential and, if not, what still needs to be done.

Materials and methods

We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [31, 32], a checklist for preparing clear and transparent accounts of systematic reviews [33], to prepare this report. Between October 2010 and October 2011 we searched the literature using PubMed (http://www.ncbi.nlm.nih.gov/pubmed/) and the Internet using Google (http://www.google.co.uk/) and Google Scholar (http://scholar.google.co.uk/) with the terms: "magnetic resonance imaging" or "MRI" or "MR" and "brain" and "databank" or "database" or "data set" and "human". The Internet search ended after two consecutive result pages provided no reference to brain image databanks. We supplemented the search by consulting the Biomedical Informatics Research Network (BIRN; http://www.birncommunity.org/resources/data/), repositories of neuroimaging resources (http://neuinfo.org/; http://www.nitrc.org/) and reference lists in previous commentaries of brain image databanks [16–19]. We first read the abstracts and/or titles produced from the search to select publications potentially describing a human brain image databank. We included databanks described in these selected publications for review if they: (1) provided publicly accessible and downloadable brain images; (2) included people aged 60 years and over; (3) described some or all of their subjects as "normal"; (4) stored structural brain images of individual subjects, but not if they stored only brain atlases or templates, i.e. averaged or combined brain images from multiple subjects, without the underlying individual subjects' images. While the latter three criteria are self-explanatory, public accessibility allows the sharing of data and results derived from these data thereby may lead to better understanding of normal brain ageing [12, 19].

We sought from all available information about each databank its: (1) purpose, (2) number of subjects, (3) number of "normal" subjects aged 60 years and over, (4) criteria for normality, (5) MR image sequences, (6) image (subject) retrieval parameters and (7) results from statistical analyses, e.g. correlation coefficients, the mean and variance of brain volumes by age, on the data contained. We estimated missing values where possible; the bases for estimates are noted in the results tables. Finally, we interfaced with databanks and/or consulted databank manuals and publications to determine how data were accessed and searched. We compiled tables summarising these results and composed individual descriptions of each databank.

Results

The literature search produced 591 publications, and we found a further 31 items through Internet and supplementary searches. Seven records were duplicates; therefore the search produced 615 individual records. We screened these

to identify 144 that potentially described a databank of human brain images. Based on criteria given in the Materials and methods section, we excluded 135 of these (Fig. 1). Particular records that did not meet inclusion criteria and the main reasons for exclusion, included:

- the Cardiovascular Health Study (CHS) database, as data were only accessible to investigators associated with the CHS study [34]
- the AddNeuroMed Study database, as data were not yet publicly accessible [35, 36]
- the BrainSCAPE database, as it was offline at the time of this study [37]
- the Whole Brain Atlas [38] and Neuroanatomical Database of Normal Japanese Brains [39, 40], as images were not downloadable
- the Allen Human Brain Atlas [41], NIH MRI Study of Normal Brain Development database [42, 43], BrainWeb Simulated MRI Volumes for Normal Brain Database [44, 45], IMAGEN project [46], Morphometry BIRN database [47] and Surface Management System Database (SumsDB) [48, 49], as they did not include subjects aged 60 years and over
- the International Stroke Database [50], Neuropsychiatric Imaging Research Laboratory (NIRL) Imaging Database [51], 1000 Functional Connectomes Project/ International Neuroimaging Data-sharing Initiative (INDI) database [52], Function BIRN Data Repository [53] and Brain Image Database (BRAID) [54, 55], as they did not describe any of their subjects as "normal"
- the BrainMap database [56, 57], BRAINnet Database [58], Brede Database [59] and Internet Brain Volume Database (IBVD) [60], as they did not contain structural brain images
- the Montreal Neurological Institute (MNI) 152 atlas [61] as it did not provide the underlying individual subjects' images.

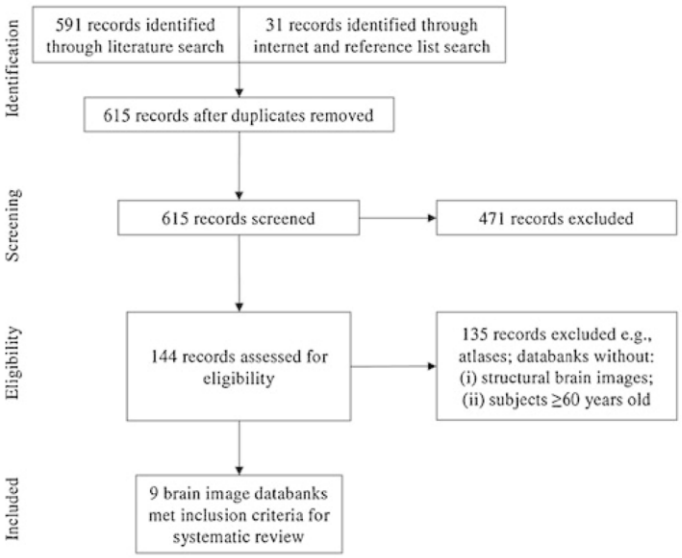


Fig. 1 Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram of systematic review phases

Summary of MR brain image databanks concerned with defining "normal" ageing brain structure

This left nine MR brain image databanks that met the criteria as follows (Table 1). All nine of these databanks reported to include "normal" subjects aged 60 years and over. However, only five databanks, ADNI, fMRIDC, OASIS cross-sectional, OASIS longitudinal and XNAT Central, included subjects fully representative of the normal ageing population according to cognitive and medical test results. Further, we could not find information on the total number of normal subjects aged ≥60 years and the criteria for normality for all subjects in the XNAT Central databank. Furthermore, many subjects were in more than one databank (fMRIDC, OASIS longitudinal, OASIS cross-sectional and XNAT Central databanks) [23, 24].

Therefore in total, according to information we could find, there were approximately 343 different individual, representative normal subjects aged ≥60 years (from the ADNI, fMRIDC and OASIS Cross-sectional databanks; Table 2). The mean age of these 343 subjects was 75.78, standard deviation (SD) 6.49 years. Many of these subjects were not accessible without application; only 98 different individual, representative normal older subjects were openly accessible. The mean age of these openly accessible subjects was 75.92, SD 8.99 years.

Regardless of whether representative metadata were available or not and discounting subject overlap, the apparent total

Table 1 MR brain image databanks concerned with defining "normal" ageing brain structure

Name [references]
1. The Alzheimer's Disease Neuroimaging Initiative (ADNI) databank [62–67]
2. Australian Imaging Biomarkers & Lifestyle Flagship Study of Ageing (AIBL) databank [22, 66–68]
3. Designed Database of MR Brain Images of Healthy Volunteers [69, 70]
4. The fMRI Data Center (fMRIDC) [71, 72]
5. Information eXtraction from Images (IXI) dataset [73–75]
6. International Consortium for Brain Mapping (ICBM) databank [16, 21, 66, 67]
7. Open Access Structural Imaging Series (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented and Demented Older Adults [24, 76]
8. OASIS: Longitudinal MRI Data in Nondemented and Demented Older Adults [23, 76]
9. The Extensible Neuroimaging Archive Toolkit (XNAT) Central [77, 78]

Table 2 Description of MR brain image databanks concerned with defining "normal" ageing brain structure

Name	No. of "normal" subjects				Age range; mean, SD in years[a]	Representative cognitive and medical test results[b]	Accessibility
	60–69	70–79	≥80 years	(Total)			
ADNI databank	0	166	63	(229)[c]	60–90; 75.8, 5.0	Yes	By application
AIBL databank	85	69	23	(177)[d]	60–100; 70.0, 7.0	No	By application
Designed Databank of MR Brain Images of Healthy Volunteers	14	1	0	(15)	60–72; 64.9, 3.1	No	Open
The fMRIDC	–	–	–	(66)[e]	60–93; 75.6, 6.7[f]	Yes	By application
IXI dataset	129	58	10	(197)	60–86; 68.1, 6.0	No	Open
ICBM databank	47	19	10	(76)[g]	60–90; –	No	By application
OASIS Cross-sectional databank	25	35	38	(98)[e]	60–94; 75.9, 9.0	Yes	Open
OASIS Longitudinal databank	23	35	28	(86)[e]	60–93; 75.8, 8.2	Yes	Open
XNAT Central	–	–	–	(–)[e]	60–94[h]; –	Yes	Open/ by application[i]

– missing value

[a] This column shows the age range, mean and standard deviation (SD) of "normal" subjects aged ≥60 years

[b] This column shows whether or not we found cognitive and medical test results, representative of the entire normal ageing population, in the databank/criteria for normality

[c] Estimated from total enrolment by age group (http://www.adni-info.org/Pdfs/ADNI_Enroll_Demographics.pdf)

[d] Estimated from age frequency distribution graph [22]

[e] Many of these subjects were part of the fMRIDC, OASIS Cross-sectional, OASIS Longitudinal and XNAT Central databanks [23, 24]

[f] Estimated from ages and sample sizes in original studies [27, 79]

[g] Estimated from exclusion by age group graph [21]

[h] This is the age range according to information that we could find

[i] Accessibility was data dependent

number of normal subjects aged ≥60 years (in the eight databanks where we could determine the number of subjects) was 944 with a mean of 118, SD 74 (range 15–229) subjects per databank. From the seven databanks where we could find or estimate age (Table 2), the mean age of subjects aged ≥60 years was 72.31, SD 6.36 years.

Table 3 shows the acquisition parameters of subject images and Table 4 the subject retrieval parameters in all of the eligible databanks. We found no databank to support subject retrieval by measures of white matter changes whether qualitative, e.g. Fazekas rating, or volumetric. No databank supported random or stratified random, i.e. random within study specific constraints such as volume of white matter changes or age, subject retrieval. Finally, no databank stored results from statistical analyses, e.g. correlation coefficients, the mean and variance of brain volumes by age, of the data it contained.

Table 3 Structural image acquisition parameters in MR brain image databanks concerned with defining "normal" ageing brain structure[a]

Name	No. of centres	T1	T2	PD	FLAIR	T2*	SWI	Tesla	T1 voxel size
ADNI databank	58	1	1	1	0	0	0	1.5, 3	1×1×1.2 mm
AIBL databank	2	1	1	1	1	0	0	3	1×1×1.2 mm
Designed Databank of MR Brain Images of Healthy Volunteers	1	1	1	0	0	0	0	3	1×1×1 mm
The fMRIDC	2	1	0	0	0	0	0	1.5	–×–×1.4 mm
IXI dataset	3	1	1	1	0	0	0	1.5, 3	0.94×0.94×1.2 mm
ICBM databank	8	1	1	1	0	0	0	1.5, 2, 3	1×1×1 mm
OASIS Cross-sectional databank	1	1	0	0	0	0	0	1.5	1×1×1.25 and 1×1×1 mm
OASIS Longitudinal databank	1	1	0	0	0	0	0	1.5	1×1×1.25 and 1×1×1 mm
XNAT Central	1	1	0	0	0	0	0	1.5	1×1×1.25 and 1×1×1 mm

PD proton density, *FLAIR* fluid-attenuated inversion recovery, *SWI* susceptibility-weighted imaging

0=image sequence was not in the databank; 1=image sequence was in the databank; – missing value

[a] For normal ageing subjects that were accessible

Table 4 Subject retrieval parameters in MR brain image databanks concerned with defining "normal" ageing brain structure

Name	Clinico-demographics	Cognitive test results	MR image acquisition parameters	Image analysis results
ADNI	1	1	1	1
AIBL	1	1	1	1
Designed Database of MR Brain Images of Healthy Volunteers	1	0	1	0
The fMRIDC	0	0	0	0
IXI dataset	1	0	1	0
ICBM	1	1	1	1
OASIS Cross-sectional	1	1	1	1
OASIS Longitudinal	1	1	1	1
XNAT Central	1	1	1	1

0=were not image retrieval parameters; 1=were image retrieval parameters

Purpose, subjects, criteria for normality and subject retrieval parameters of each databank

ADNI databank

The ADNI was set up to determine which combination of neuroimaging, cerebral spinal fluid (CSF) and blood biomarkers provides the earliest and most accurate diagnosis and expected course of Alzheimer's disease (AD). Housed at the Laboratory of Neuroimaging (LONI) Image Data Archive (IDA), the ADNI databank contained serial MR brain images, separated by 6–12 months over 2–3 years, from approximately 229 normal, 398 mild cognitive impairment (MCI) and 192 AD subjects aged 55–90 years (normal subjects were aged 70-90 years; Table 2). Normal subjects may have had some medical problems common in ageing. The criteria for normality also included results from a battery of cognitive tests including the American National Adult Reading Test (ANART), Clinical Dementia Rating (CDR) and Mini Mental State Examination (MMSE), which were in the databank and subject retrieval parameters. Image analysis results were in the databank but we did not find statistical results. The LONI IDA also supported subject retrieval by a range of clinico-demographic, image acquisition and image analysis parameters.

AIBL databank

The AIBL study was designed to understand the pathological features and early clinical manifestation of AD, improve the diagnosis of AD, and identify diet and lifestyle factors that are significant in the development of AD. Also housed at the LONI IDA, the AIBL databank contained serial MR brain images acquired from 177 normal, 57 MCI and 53 AD subjects aged 60–100 years. Although medical and cognitive test results were part of the criteria for normality, the normal subjects were not representative of the normal ageing population because they were preferentially selected as APOE ε4 allele carriers [68].

Designed Database of MR Brain Images of Healthy Volunteers

The Designed Database of MR Brain Images of Healthy Volunteers was created to assess the effects of healthy ageing on brain structure and provide references for the assessment of disease. It contained MR brain images from 100 normal subjects aged 18–72 years (15 subjects were aged ≥60 years; Table 2). These subjects had "no history of diabetes, hypertension, head trauma, psychiatric disease, or other symptoms or history likely to affect the brain" and therefore may not have been representative of the entire normal ageing population. We did not find cognitive test results to be within the criteria for normality. Subject demographics age, gender, race and handedness were in the database, and these could be used to retrieve subjects from the database.

The fMRIDC

The fMRIDC was created to allow the functional magnetic resonance imaging (fMRI) research community to validate methods and hypotheses and perform meta-analyses of a large number of peer-reviewed studies. It stored structural MR brain images as well as fMRI data from these studies. We did not have access to the overall composition of the databank, e.g. we could not determine the total number of subjects, but found 66 normal subjects aged ≥60 years. These subjects had no history of stroke, heart attack or psychiatric disorder but had medical characteristics common in ageing, e.g. hypertension and arthritis. Fifty of these subjects were also in the OASIS and XNAT Central databanks. Subjects could be

retrieved by study title, author, keywords, abstract or "special collections" (selected novel data sets) but not directly by subject clinico-demographic, imaging or cognitive parameters.

IXI dataset

The IXI dataset was acquired to develop computer-aided diagnostics of MR brain images. It contained MR brain images from 593 normal, healthy subjects aged 19–86 years (197 subjects were aged ≥60 years; Table 2). We did not find the criteria for normality, and medical and cognitive test results were not in the data set. Subject demographics such as age, gender, weight, ethnicity and qualification were in the data set, and these could be used to retrieve subjects from the data set.

ICBM databank

The ICBM study will develop a probabilistic atlas and reference system for the normal human brain throughout the lifespan [16]. The ICBM databank, also housed at the LONI IDA, had 851 normal subjects aged 18–90 years (approximately 76 subjects were aged ≥60 years; Table 2). The criteria for normality included results from several medical and cognitive tests such as the MMSE. As the criteria for normality were the same regardless of age, the older subjects may not have been representative of the entire normal ageing population, e.g. subjects with any prescription medications (with some exceptions such as antibiotics or non-steroidal anti-inflammatories) or hypertension were excluded regardless of age.

OASIS: Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults

The OASIS cross-sectional databank was created to provide the data needed, for example, for widespread study of ageing and dementia and to develop new MR brain image analysis techniques. It contained MR brain images from 316 normal subjects and 100 dementia sufferers aged 18–96 years (98 normal subjects were aged ≥60 years; Table 2), including older subjects with hypertension and treated diabetes. The criteria for normality also included MMSE score and CDR. The databank supported subject retrieval by a range of clinico-demographics, cognitive test results (CDR and MMSE), imaging acquisition parameters and parameters derived from analyses of brain images: atlas scaling factor, estimated total intracranial volume, whole brain volume and normalised whole brain volume, i.e. proportion of whole brain volume in total intracranial volume. Although these image analysis results were in the databank, statistical results were not.

OASIS: Longitudinal MRI Data in Nondemented and Demented Older Adults

The OASIS longitudinal databank was created for reasons similar to the OASIS cross-sectional databank. It contained serial MR brain images, acquired over two or more sessions and separated by at least 1 year, from 86 normal subjects (14 of which later converted to dementia) and 64 dementia sufferers aged 60–96 years. Many of the older subjects in the OASIS cross-sectional databank were also in this longitudinal databank but were assigned new subject identifiers. The criteria for normality and subject retrieval parameters were the same as in the OASIS cross-sectional databank.

The Extensible Neuroimaging Archive Toolkit (XNAT) Central databank

The XNAT Central databank was designed to allow secure and quality-controlled data (medical image and metadata) sharing among local colleagues, external collaborators and the broader neuroscience community. It contained over 3,000 subjects from approximately 200 medical imaging studies, including the OASIS data. In addition to the OASIS subjects we found nine subjects (with brain images) aged ≥60 years; however, these subjects were not normal (they were neurosurgery patients). We did not have access to the remaining subjects or criteria for normality. Subjects could be retrieved by the parameters in the OASIS cross-sectional databank among others, including medical test results and regional brain volumes.

Discussion

We systematically reviewed published MR brain image databanks with structural brain images of "normal" older people (aged ≥60 years). Amongst nine databanks that met the inclusion criteria, concerned with defining "normal" ageing brain structure, there appeared to be 944 normal subjects aged 60 years and over. However, after adjusting for the many subjects who were in more than one databank and those who did not have metadata (cognitive and medical test results) representative of the entire normal ageing population, there were 343 normal subjects aged 60 years and over, only 98 of whom were openly accessible. While lack of open access is not necessarily a criticism (MR imaging data are expensive to acquire and easily misused), application reviewer bias may restrict investigators with intentions not aligned with those of the original studies. The high variation in structure of the normal ageing brain [1, 4, 5] and inconsistency of causal inferences [2, 3, 6, 14] indicate that this number of subjects (most of whom some investigators may not be able to access) is too few to effectively characterise this variation.

The criteria for normality in many of these databanks may not have fully represented the clinical characteristics of normal ageing. For example, the Designed Database of MR Brain Images of Healthy Volunteers and ICBM criteria for normality were the same across the lifespan [21, 70] and thus may not have been fully representative of the normal ageing population's clinical characteristics such as the increasing proportion with prescription medications and hypertension. When a population has a mean of approximately three prescriptions and approximately 50% have medically diagnosed hypertension [1, 21, 25], it would be reasonable to conclude that subjects with these characteristics are at least equally "normal" to subjects without and both should be included in normal ageing brain image databanks. Thus, representing the entire normal ageing population that includes "successful" and "usual" ageing individuals [28]. This is particularly true when considering that these subjects may be used as controls for study groups, e.g. dementia subjects, that have similar clinical characteristics [23].

Magnetic resonance brain image databanks with representative normal older subjects have led to many publications: over 200 publications from the ADNI databank and the OASIS cross-sectional databank has been cited by over 100 publications [80, 81]. However, these databanks provided a limited range of image sequences [63]; the openly accessible images from representative normal older subjects were only T1-weighted [24]. To support better understanding of normal ageing brain structure, databanks should include, for example, T2* and FLAIR images as well as the commonly used T1- and T2-weighted images [20, 24, 63].

Almost all databanks supported image (subject) retrieval by clinico-demographic and imaging acquisition parameters but few (LONI IDA, OASIS and XNAT Central) supported subject retrieval directly by medical and cognitive test results and parameters derived from image analyses, e.g. brain volumes. In particular, we did not find any databank to support subject retrieval by white matter changes that, if prevalent in a normal ageing control group, could skew cognitive measures [82]. Therefore, it may be difficult to match or differentiate databank subjects (to study groups) on parameters, e.g. age, brain structure, MMSE, head size and blood pressure, that could undesirably affect experimental results when using them as controls. Moreover, no databank supported random subject retrieval; therefore, manual selection bias and errors may occur in studies that use these subjects as controls.

Further to storing subjects, databanks that store results from statistical analyses of normal brain images and metadata may facilitate better understanding of normal ageing brain structure [1, 11–13, 19], e.g. results from statistical analyses (such as the mean and variance of brain volumes by age) can be stored, tested and validated with analyses of further data. The storage of statistical results is required owing to the unreliability of causal inferences from single studies [11–13] and particularly so given the inconsistencies in reports of the progression of atrophy in normal ageing and correlations of cognition and brain volumes [2, 3, 6, 14]. However, we found no databank that had these results. Although not yet included, the ICBM databank plans to include probabilistic atlases (maps) of the variation of normal whole and regional brain structures from 7,000 subjects throughout the adult lifespan [16]. However, according to a recent progress report of this work [21], the number of older subjects (aged ≥60 years) to be included in these atlases will be limited (estimated 581, from subject proportions in the progress report, if the 7,000 subject target is reached). The Internet Brain Volume Database (IBVD) [60], excluded from our review as it did not contain structural brain images, contained brain volumes from different studies with different parameters and criteria for normality, and hence may not effectively describe normal ageing brain structure variation. Therefore, there is a current lack of easily accessible statistical results, e.g. the distribution of normal ageing brain volumes, and atlases (from representative normal ageing subjects) that may facilitate better understanding of normal ageing brain structure and the required earlier and more accurate diagnoses of dementia and related disorders [1, 9, 11–13]. To have such a great effect, e.g. so as to be used in large clinical trials that may, for example, test the effectiveness of dementia preventing treatments according to patient positions in the distribution of normal ageing brain volumes, these results and atlases should be openly accessible in a brain image databank [19].

The strengths of this study include use of established systematic review criteria, exhaustive search of printed and online materials, and structured evaluation of databanks according to prespecified criteria. The study limitations include difficulty in searching for databanks and our search process, although we used multiple overlapping techniques, may not have found every databank with structural brain images from "normal" older subjects. In particular, we were not able to comment on databanks that were not publicly accessible. Further, we used available publications, user guides and fact sheets, and interfaced with databanks when possible. However, we may not have found all relevant information so may, inadvertently, not have justly described the databanks we did find, for which we apologise. These limitations notwithstanding, we have considered all of the leading published brain image databanks [19], and others not previously reviewed, that are concerned with the variation in structure of the "normal" ageing brain.

According to our review, brain image databanks with normal older subjects have the potential to facilitate better understanding of the normal ageing brain structure. This understanding is ever more important as cases of age-

related neurological disorders grow with the average life expectancy, which is now on average in the 9th decade for western countries. However, the total number of openly accessible subjects fully representative of the normal ageing population's clinical characteristics (approximately 98 different individual subjects aged ≥60 years) in existing brain image databanks is too limited at present to inform the true variation in normal ageing brain structure.

Databanks should include subjects thoroughly tested to show no cognitive or other debilitating disorder but with clinical characteristics that are common in ageing, e.g. prescription medications, diabetes, hypertension and arthritis [1, 21–27]. As long as these characteristics are carefully documented and searchable, this will allow others to draw subjects appropriate for their study group and/or representative of the wider population. To avoid bias and errors in brain imaging studies that use these subjects as controls, databanks should support random and stratified random subject retrieval. Multiple image sequences, including the T2, T2* and FLAIR, as well as T1-weighted sequences, that are routinely used in diagnosis, are required to define the structure of the normal and abnormal ageing brain. Finally, results of statistical analyses and atlases defining normal ageing brain structure variation should be included in databanks to provide a reference for others to test. With further data and analyses these results and atlases will be validated or evolve to facilitate better understanding of normal ageing brain structure. In turn, this understanding may lead to earlier and more accurate diagnoses of disorders such as dementia and facilitate clinical trials of new treatments.

Acknowledgements This work was carried out in The University of Edinburgh Brain Research Imaging Centre (BRIC; http://www.bric.ed.ac.uk/) and the University of Aberdeen Biomedical Imaging Centre (http://www.abdn.ac.uk/ims/imaging/)—both centres are part of the Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) collaboration (http://www.sinapse.ac.uk/) that is funded by the Scottish Funding Council, Scottish Executive Chief Scientist Office, and the six collaborator Universities—and in Toshiba Medical Visualisation Systems Europe (TMVSE; http://www.tmvse.com/). We thank the funders of this work as follows. Prof. Joanna M. Wardlaw was funded by the Scottish Funding Council and Scottish Executive Chief Scientist Office through the SINAPSE collaboration; David Alexander Dickie was funded by a SINAPSE industrial collaboration (SPIRIT) PhD scholarship with TMVSE, a Medical Research Council (MRC) scholarship, and the Tony Watson Scholarship bequest to The University of Edinburgh; Dr Dominic E. Job was funded by Wellcome Trust Grant 007393/Z/05/Z; Dr Trevor S. Ahearn was funded by SINAPSE and the University of Aberdeen; Dr Roger T. Staff was funded by NHS Grampian; Dr Alison D. Murray was funded by NHS Grampian via the University of Aberdeen.

References

1. Farrell C, Chappell F, Armitage P, Keston P, MacLullich A, Shenkin S, Wardlaw JM (2009) Development and initial testing of normal reference MR images for the brain at ages 65–70 and 75–80 years. Eur Radiol 19:177–183
2. Fotenos AF, Snyder A, Girton L, Morris J, Buckner R (2005) Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. Neurology 64:1032–1039
3. Good CD, Johnsrude IS, Ashburner J, Henson RNA, Friston KJ, Frackowiak RSJ (2001) A voxel-based morphometric study of ageing in 465 normal adult human brains. NeuroImage 14:21–36
4. Manolio TA, Kronmal RA, Burke GL, Poirier V, O'Leary DH, Gardin JM, Fried LP, Steinberg EP, Bryan RN (1994) Magnetic resonance abnormalities and cardiovascular disease in older adults. The Cardiovascular Health Study. Stroke 25:318–327
5. Resnick SM, Pham DL, Kraut MA, Zonderman AB, Davatzikos C (2003) Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. J Neurosci 23:3295–3301
6. Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW (2003) Mapping cortical change across the human life span. Nat Neurosci 6:309–315
7. Thompson PM, Hayashi KM, de Zubicaray GI, Janke AL, Rose SE, Semple J, Herman D, Hong MS, Dittmer SS, Doddrell DM, Toga AW (2003) Dynamics of gray matter loss in Alzheimer's disease. J Neurosci 23:994–1005
8. Luengo-Fernandez R, Leal J, Gray A (2010) Dementia 2010: The economic burden of dementia and associated research funding in the United Kingdom. University of Oxford for the Alzheimer's Research Trust
9. Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. Physiol Rev 81:741–766
10. Department of Health (2009) Living well with dementia: A National Dementia Strategy. London
11. Cohen J (1994) The earth is round (p<.05). Am Psychol 49:997–1003
12. Freedman D (2010) Statistical Models and Causal Inference: A Dialogue with the Social Sciences. Cambridge University Press, Cambridge
13. Meehl PE (1978) Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. J Consult Clin Psychol 46:806–834
14. Salthouse TA (2011) Neuroanatomical substrates of age-related cognitive decline. Psychol Bull 137:753–784
15. Insel TR, Volkow ND, Landis SC, Li TK, Battey JF, Sieving P (2004) Limits to growth: why neuroscience needs large-scale science. Nat Neurosci 7:426–427
16. Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B, Holmes C, Collins L, Thompson P, MacDonald D, Iacoboni M, Schormann T, Amunts K, Palomero-Gallagher N, Geyer S, Parsons L, Narr K, Kabani N, Le Goualher G, Boomsma D, Cannon T, Kawashima R, Mazoyer B (2001) A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos Trans R Soc Lond B Biol Sci 356:1293–1322
17. Toga AW (2002) Neuroimage databases: the good, the bad and the ugly. Nat Rev Neurosci 3:302–309
18. Toga AW, Thompson PM, Mori S, Amunts K, Zilles K (2006) Towards multimodal atlases of the human brain. Nat Rev Neurosci 7:952–966
19. Van Horn JD, Toga AW (2009) Is it time to re-prioritize neuroimaging databases and digital repositories? NeuroImage 47:1720–1734
20. Wardlaw JM, Bastin ME, Valdés Hernández MC, Muñoz Maniega S, Royle NA, Morris Z, Clayden JD, Sandeman EM, Eadie E, Murray C, Starr JM, Deary IJ (2011) Brain ageing, cognition in youth and old age, and vascular disease in the Lothian Birth Cohort 1936: rationale, design and methodology of the imaging protocol. Int J Stroke 6:547–559

21. Mazziotta JC, Woods R, Iacoboni M, Sicotte N, Yaden K, Tran M, Bean C, Kaplan J, Toga AW (2009) The myth of the normal, average human brain−The ICBM experience: (1) Subject screening and eligibility. NeuroImage 44:914–922
22. Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P (2009) The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 21:672–687
23. Marcus DS, Fotenos AF, Csernansky JG, Morris JC, Buckner RL (2010) Open access series of imaging studies (OASIS): Longitudinal MRI data in nondemented and demented older adults. J Cogn Neurosci 22:2677–2684
24. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 19:1498–1507
25. DeCarli C, Massaro J, Harvey D, Hald J, Tullberg M, Au R, Beiser A, D'Agostino R, Wolf PA (2005) Measures of brain morphology and infarction in the Framingham Heart Study: establishing what is normal. Neurobiol Aging 26:491–510
26. Jernigan TL, Archibald SL, Fennema-Notestine C, Gamst AC, Stout JC, Bonner J, Hesselink JR (2001) Effects of age on tissues and regions of the cerebrum and cerebellum. Neurobiol Aging 22:581–594
27. Grady CL, Springer MV, Hongwanishkul D, McIntosh AR, Winocur G (2006) Age-related changes in brain activity across the adult lifespan. J Cogn Neurosci 18:227–241
28. Barkhof F, Fox NC, Bastos-Leite AJ, Scheltens P (2011) Neuroimaging in Dementia. Springer, Berlin Heidelberg
29. Folstein M, Folstein S, McHugh P (1975) "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 12:189–198
30. Morris JC (1993) The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43:2412–2414
31. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine 6:e1000097
32. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) The PRISMA Statement. Available: http://www.prisma-statement.org/statement.htm. Accessed 31 May 2011
33. EQUATOR Network (2011) About EQUATOR. Available: http://www.equator-network.org/about-equator/. Accessed 31 May 2011
34. The Cardiovascular Health Study (2010) Data Distribution Policy. Available: http://www.chs-nhlbi.org/CHS_DistribPolicy.htm. Accessed 31 May 2011
35. The AddNeuroMed Study (2011) Data Access. Available: http://www.innomed-addneuromed.com/index.cfm?PID=108. Accessed 07 September 2011
36. Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, Koszewska I, Wahlund LO, Soininen H, Lovestone S, Evans A, Spenger C (2011) The AddNeuroMed framework for multi centre MRI assessment of Alzheimer's disease: experience from the first 24 months. Int J Geriatr Psychiatry 26:75–82
37. Neuroinformatics Research Group (2011) BrainSCAPE. Available: http://nrg.wustl.edu/project/data-sharing/. Accessed 10 October 2011
38. Johnson KA, Becker JA (1999) The Whole Brain Atlas. Available: http://www.med.harvard.edu/AANLIB/home.html. Accessed 31 May 2011
39. Sato K, Taki Y, Fukuda H, Kawashima R (2003) Neuroanatomical database of normal Japanese brains. Neural Netw 16:1301–1310
40. Sato K, Taki Y, Fukuda H, Kawashima R (2003) Japanese Reference Brains. Available: http://www.idac.tohoku.ac.jp/JHBP/. Accessed 31 May 2011
41. Allen Institute for Brain Science (2011) Allen Human Brain Atlas. Available: http://human.brain-map.org/mri_viewers/data. Accessed 31 May 2011
42. Evans AC (2006) The NIH MRI study of normal brain development. NeuroImage 30:184–202
43. Evans AC (2006) The NIH MRI Study of Normal Brain Development Database. Available: https://nihpd.crbs.ucsd.edu/nihpd/info/index.html. Accessed 31 May 2011
44. Cocosco CA, Kollokian V, Remi KSK, Pike GB, Evans AC (1997) Brainweb: Online interface to a 3D MRI simulated brain database. NeuroImage 5:s425
45. McConnell Brain Imaging Centre of the Montreal Neurological Institute (2004) BrainWeb: Simulated MRI Volumes for Normal Brain Database. Available: http://mouldy.bic.mni.mcgill.ca/brainweb/selection_normal.html. Accessed 31 May 2011
46. The IMAGEN Consortium (2011) Imagen Europe−6th Framework Project. Available: http://www.imagen-europe.com/. Accessed 31 May 2011
47. Biomedical Informatics Research Network (BIRN) (2009) Morphometry BIRN Multi-site Multi-session Structural MRI Data. Available: http://www.birncommunity.org/data-catalog/morphometry-birn-multi-site-multi-session-structural-mri-data/. Accessed 10 October 2011
48. Van Essen Lab (2001) Surface Management System Database. Available: http://sumsdb.wustl.edu:8081/sums/index.jsp. Accessed 31 May 2011
49. Dickson J, Drury H, Van Essen DC (2001) The surface management system (SuMS) database: a surface–based database to aid cortical surface reconstruction, visualization and analysis. Philos Trans R Soc Lond B Biol Sci 356:1277–1292
50. Sorensen AG, Wu O (2010) International Stroke Database. Available: http://www.strokedatabase.org/. Accessed 31 May 2011
51. Neuropsychiatric Imaging Research Laboratory (2009) NIRL Imaging Database. Available: http://nirlarc.duhs.duke.edu/nirle/. Accessed 31 May 2011
52. Milham M, Buckner RL, Castellanos FX, Margulies D, Zang Y, Mennes M, Gutman D, Bangaru S, Craddock C, LaConte S, Mostofsky S, Villringer A (2011) 1000 Functional Connectomes Project. Available: http://fcon_1000.projects.nitrc.org/index.html. Accessed October 10 2011
53. Biomedical Informatics Research Network (BIRN) (2011) Function BIRN Data Repository. Available: http://fbirnbdr.nbirn.net:8080/BDR/index.jsp. Accessed 07 October 2011
54. Letovsky SI, Whitehead S, Paik CH, Miller GA, Gerber J, Herskovits EH, Fulton TK, Bryan RN (1998) A brain image database for structure/function analysis. AJNR Am J Neuroradiol 19:1869–1877
55. Department of Radiology University of Pennsylvania (2008) Brain Image Database (BRAID). Available: http://www.rad.upenn.edu/sbia/braid/. Accessed 31 May 2011
56. Fox PT, Lancaster JL (2002) Mapping context and content: the BrainMap model. Nat Rev Neurosci 3:319–321
57. Research Imaging Institute UTHSCSA (2010) BrainMap Database. Available: http://brainmap.org/index.html. Accessed 31 May 2011
58. BRAINnet Foundation (2009) BRAINnet Database. Available: http://www.brainnet.net/what-data-are-available/mri-fmri-and-dti/. Accessed 31 May 2011
59. Technical University of Denmark Informatics (2009) Brede Database. Available: http://neuro.imm.dtu.dk/services/jerne/brede/. Accessed 31 May 2011
60. The Center for Morphometric Analysis MGH HMS (2002) Internet Brain Volume Database. Available: http://www.cma.mgh.harvard.edu/ibvd/. Accessed 31 May 2011
61. McConnell Brain Imaging Centre of the Montreal Neurological Institute (2010) Atlases. Available: http://www.bic.mni.mcgill.ca/ServicesAtlases/HomePage. Accessed 31 May 2011

62. Petersen R, Aisen P, Beckett L, Donohue M, Gamst A, Harvey D, Jack C, Jagust W, Shaw L, Toga A (2010) Alzheimer's Disease Neuroimaging Initiative (ADNI). Neurology 74:201
63. Jack CR Jr, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ (2008) The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 27:685–691
64. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L (2005) Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement 1:55–66
65. Laboratory of Neuro Imaging (2011) Alzheimer's Disease Neuroimaging Initiative (ADNI). Available: http://adni.loni.ucla.edu/. Accessed 31 May 2011
66. Laboratory of Neuro Imaging (2011) LONI Image Data Archive (IDA)−Data Access. Available: https://ida.loni.ucla.edu/services/Menu/IdaData.jsp?page=DATA&subPage=AVAILABLE_DATA. Accessed 12 May 2011
67. Toga AW (2009) LONI Image Data Archive User Manual. Laboratory Of Neuro Imaging, UCLA
68. Ellis KA, Rowe CC, Villemagne VL, Martins RN, Masters CL, Salvado O, Szoeke C, Ames D (2010) Addressing population aging and Alzheimer's disease through the Australian imaging biomarkers and lifestyle study: Collaboration with the Alzheimer's disease neuroimaging initiative. Alzheimers Dement 6:291–296
69. Mortamet B, Zeng D, Gerig G, Prastawa M, Bullitt E (2005) Effects of healthy aging measured by intracranial compartment volumes using a designed MR brain database. Medical Image Computing and Computer-Assisted Intervention (MICCAI) 3749:383–391
70. Bullitt E, Smith JK, Lin W (2010) Designed Database of MR Brain Images of Healthy Volunteers. Available: http://www.insight-journal.org/midas/community/view/21. Accessed 31 May 2011
71. Van Horn JD, Grethe JS, Kostelec P, Woodward JB, Aslam JA, Rus D, Rockmore D, Gazzaniga MS (2001) The functional magnetic resonance imaging data center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. Philos Trans R Soc Lond B Biol Sci 356:1323–1339
72. Van Horn JD, Grethe JS, Kostelec P, Woodward JB, Aslam JA, Rus D, Rockmore D, Gazzaniga MS (2007) The fMRI Data Center. Available: http://www.fmridc.org/f/fmridc/index.html. Accessed 31 May 2011
73. Biomedical Image Analysis Group Imperial College London (2010) Information eXtraction from Images (IXI) dataset. Available: http://www.brain-development.org/. Accessed 31 May 2011
74. Hill DLG, Hawkes D, Williams S (2010) Information eXtraction from Images (IXI): Details of Grant. Available: http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=GR/S21533/02. Accessed 31 May 2011
75. Rowland A, Burns M, Hartkens T, Hajnal JV, Rueckert D, Hill DLG (2004) Information extraction from images (IXI): Image processing workflows using a grid enabled image database. Distributed Databases in Medical Image Computing Workshop MICCAI, Rennes, France
76. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open Access Series of Imaging Studies (OASIS). Available: http://www.oasis-brains.org/. Accessed 31 May 2011
77. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL (2007) The extensible neuroimaging archive toolkit. Neuroinformatics 5:11–33
78. Marcus DS, Olsen TR, Ramaratnam M, Buckner RL (2011) XNAT Central. Available: http://central.xnat.org/. Accessed 31 May 2011
79. Head D, Snyder AZ, Girton LE, Morris JC, Buckner RL (2005) Frontal-hippocampal double dissociation between normal aging and Alzheimer's disease. Cereb Cortex 15:732–739
80. ADNI (2011) ADNI Publications. Available: http://www.adni-info.org/Scientists/ADNIScientistsHome/ADNIPublications.aspx. Accessed 26 June 2011
81. Google Scholar (2011) Search within articles citing Marcus: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Available: http://scholar.google.co.uk/scholar?cluster=14063688880671780453&hl=en&as_sdt=2005&sciodt=1,5. Accessed 2 September 2011
82. Breteler M, Van Swieten J, Bots M, Grobbee D, Claus J, Van Den Hout J, Van Harskamp F, Tanghe H, De Jong P, Van Gijn J (1994) Cerebral white matter lesions, vascular risk factors, and cognitive function in a population-based study. Neurology 44:1246–1246

PLOS ONE

Variance in brain volume with advancing age: implications for defining the limits of normality

--Manuscript Draft--

Manuscript Number:	PONE-D-13-40278R1
Article Type:	Research Article
Full Title:	Variance in brain volume with advancing age: implications for defining the limits of normality
Short Title:	Normal ageing brain volume variance and limits
Corresponding Author:	David Alexander Dickie The University of Edinburgh Edinburgh, UNITED KINGDOM
Keywords:	brain structure; Ageing; alzheimer's disease; magnetic resonance imaging; regression analysis; percentile ranks
Abstract:	Background: Statistical models of normal ageing brain tissue volumes may support earlier diagnosis of increasingly common, yet still fatal, neurodegenerative diseases. For example, the statistically defined distribution of normal ageing brain tissue volumes may be used as a reference to assess patient volumes. To date, such models were often derived from mean values which were assumed to represent the distributions and boundaries, i.e. percentile ranks, of brain tissue volume. Since it was previously unknown, the objective of the present study was to determine if this assumption was robust, i.e. whether regression models derived from mean values accurately represented the distributions and boundaries of brain tissue volume at older ages. Materials and Methods: We acquired T1-w magnetic resonance (MR) brain images of 227 normal and 219 Alzheimer's disease (AD) subjects, all aged 55-89 years, from publicly available databanks. Using nonlinear regression within both samples, we compared mean and percentile rank estimates of whole brain tissue volume by age. Results: In both the normal and AD sample, mean regression estimates of brain tissue volume often did not accurately represent percentile rank estimates (errors=-74% to 75%). In the normal sample, mean estimates generally underestimated differences in brain volume at percentile ranks below the mean. Conversely, in the AD sample, mean estimates generally underestimated differences in brain volume at percentile ranks above the mean. Differences between ages at the 5th percentile rank of normal subjects were 38.7% greater than mean differences in the AD subjects. Conclusions: While more data are required to make true population inferences, our results indicate that mean regression estimates may not accurately represent the distributions of ageing brain tissue volumes. This suggests that percentile rank estimates will be required to robustly define the limits of brain tissue volume in normal ageing and neurodegenerative disease.
Order of Authors:	David Alexander Dickie
	Dominic E Job
	David Rodriguez Gonzalez
	Susan D Shenkin
	Trevor S Ahearn
	Alison D Murray
	Joanna M Wardlaw
Suggested Reviewers:	Nick Fox University College London n.fox@ucl.ac.uk Professor Fox's research interests and expertise are closely related to this work.

	Chris Rorden University of South Carolina rorden@sc.edu Professor Rorden's research interests and expertise are closely related to this work.
Opposed Reviewers:	
Response to Reviewers:	Response to reviewers of submission, “Variance in brain volume with advancing age: implications for defining the limits of normality” Academic Editor comment 1 We note that you received funding from a commercial source Toshiba Medical Visualisation Systems Europe. Please respond in the cover letter to declare this commercial funder, along with any other relevant declarations relating to employment, consultancy, patents, products in development or marketed products etc) and if true, you should also confirm that this does not alter your adherence to all the PLOS ONE policies on sharing data and materials, as detailed online in our guide for authors http://www.PLOSone.org/static/editorial.action#competing by including the following statement: "This does not alter our adherence to all the PLOS ONE policies on sharing data and materials.” Response to Academic Editor comment 1 We have responded in the cover letter declaring the role of Toshiba Medical Visualisation Systems Europe. This relationship does not alter our adherence to PLOS ONE policies and we have included the requested statement. Reviewer #1 comment 1 My main concern with the manuscript is the methodology. Specifically, to my best understanding, the authors have used linear regression to fit the mean differences in GM/WM/CSF between age groups. Beta coefficients from the regression for mean and different percentile ranks (defined in Eq. 2) are then used to calculate proportional error (Eq. 3). Since the populations included in this study ranges from 55 to 89 years of age, the regression is assuming a linear relationship between age and brain tissue/CSF volume, which does not seem to be a solid assumption. Can the authors point to studies in the past that validated this assumption? If not, then the linear regression performed for the mean and percentile ranks are all in question as to their validity in characterizing the aging effects. Response to Reviewer #1 comment 1 We agree that, although applied in previous studies (e.g. Good et al., 2001, ref [30]), the linearity assumption (and subsequent use of linear regression) was questionable. Therefore, we revised the analysis to initially perform linear regression from which we then created residual plots (actual minus linear regression predicted brain tissue volume). This is a statistical diagnostic of nonlinearity. As we found a systematic pattern in this plot (revised figure 1), we performed nonlinear (rather than linear) regression to define mean and percentile rank estimates of brain tissue volume by age (revised figures 2 and 4, and revised tables 4 and 5). Our main finding remained the same after use of nonlinear regression: mean estimates often did not well approximate percentile rank estimates. Reviewer #1 comment 2 Another concern with the study is the relatively small number of subjects in each age group for such an ambitious goal. For example, there were about 10 to 20 subjects in each of the normal group in 3 age ranges. With this number of subjects, it is not hard to imagine that the percentile ranks could be very noisy and could be unintentionally skewed, which can explain the large proportional errors listed in the results. Response to Reviewer #1 comment 2 We agree that sample size was creating an element of noise in percentile ranks. To address this we combined the two normal subject groups and analysed the combined

	sample (n=227). Further, we added Alzheimer's disease (AD) subjects from the Open Access Series of Imaging Studies (OASIS; www.oasis-brains.org) to those that we already had from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We then analysed this combined AD sample of n=219. According to a systematic review published at the time of our analysis (ref [5] in our manuscript), OASIS and ADNI were the only public sources of structural brain images with medical and cognitive metadata that represented the characteristics of "normal" ageing. We agree that even with these additional data, our samples are still insufficient to define true population variance and limits. This is not our intention at this stage. We only intend our results to show the principle that, at least in presently available data, mean regression estimates may not best represent the distributions and boundaries of ageing brain tissue volumes. Instead, these are measures that can be calculated relatively simply via percentile rank analysis, independent of mean estimates. While investigating noise in percentile ranks, it was apparent that the misclassification of hypointense white matter regions as grey matter (discussed in section, "2.2 MR brain image acquisition and processing") was also a contributory factor. Although incorrectly classified as GM, these regions were still correctly classified as tissue, i.e. not cerebrospinal fluid. So to address the noise from these misclassifications we combined WM and GM volumes to create "whole brain tissue" regression models, i.e. (GM+WM)/(GM+WM+CSF). With the increased sample sizes and whole brain models, we feel that we have removed as much noise as possible at present. Even after removal of this noise, mean estimates often did not well approximate percentile rank estimates. Reviewer #1 comment 3 In page 4, 2nd paragraph, "although it is generally assumed, it is not known whether mean estimates of brain volume across age are actually equal to "percentile rank" estimates." This sentence is vague and confusing. Did anybody suggest that mean estimates of brain volume should be EQUAL to percentile rank estimates? Response to Reviewer #1 comment 3 We agree and apologise that this sentence was not clear. We have replaced this with the following sentence at the start of the second paragraph on page 4, "Specifically, it is not known whether mean (parametric) estimates of ageing brain volumes approximate percentile rank estimates. In parametric regression models it is implicitly assumed that mean estimates approximate percentile rank estimates [6]."
Additional Information:	
Question	**Response**
Competing Interest For yourself and on behalf of all the authors of this manuscript, please declare below any competing interests as described in the "PLoS Policy on Declaration and Evaluation of Competing Interests." You are responsible for recognizing and disclosing on behalf of all authors any competing interest that could be perceived to bias their work, acknowledging all financial support and any other relevant financial or competing interests. If no competing interests exist, enter: "The authors have declared that no competing interests exist." If you have competing interests to declare, please fill out the text box completing the following statement: "I	The authors have declared that no competing interests exist.

have read the journal's policy and have the following conflicts" * typeset	
Financial Disclosure Describe the sources of funding that have supported the work. Please include relevant grant numbers and the URL of any funder's website. Please also include this sentence: "The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript." If this statement is not correct, you must describe the role of any sponsors or funders and amend the aforementioned sentence as needed. * typeset	This work was carried out in The University of Edinburgh Brain Research Imaging Centre (BRIC; http://www.bric.ed.ac.uk/) and the University of Aberdeen Biomedical Imaging Centre (http://www.abdn.ac.uk/ims/imaging/). Both centres are part of the Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) collaboration (http://www.sinapse.ac.uk/), funded by the Scottish Funding Council, Scottish Executive Chief Scientist Office, and the six collaborator Universities. Professor Joanna M. Wardlaw was funded by the Scottish Funding Council and Scottish Executive Chief Scientist Office through the SINAPSE collaboration. David Alexander Dickie was funded by a SINAPSE industrial collaboration (SPIRIT) PhD scholarship with Toshiba Medical Visualisation Systems Europe (http://www.tmvse.com/), a Medical Research Council (MRC) scholarship, and the Tony Watson Scholarship bequest to The University of Edinburgh. Dr Dominic E. Job was funded by Wellcome Trust Grant 007393/Z/05/Z. Dr Trevor S. Ahearn was funded by SINAPSE and the University of Aberdeen. Professor Alison D. Murray was funded by NHS Grampian via the University of Aberdeen. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.Toshiba Medical Visualisation Systems Europe (TMVSE) provided support during David Alexander Dickie's PhD training. This included provision of an industrial placement onsite with TMVSE and partial funding to support conference attendance. TMVSE did not fund the PhD stipend or fees, participate in analysis or writing of our submission and this relationship does not alter our adherence to all the PLOS ONE policies on sharing data and materials.
Ethics Statement All research involving human participants must have been approved by the authors' institutional review board or equivalent committee(s) and that board must be named by the authors in the manuscript. For research involving human participants, informed consent must have been obtained (or the reason for lack of consent explained, e.g. the data were analyzed anonymously) and all clinical investigation must have been conducted according to the principles expressed in the Declaration of Helsinki. Authors should submit a statement from their ethics committee or institutional review board indicating the approval of the research. We also encourage authors to submit a sample of a patient consent form and may require submission of completed forms on particular occasions. All animal work must have been conducted according to relevant national and international guidelines. In accordance with the recommendations of the Weatherall report, "The use of non-human primates in research" we specifically require authors to include details of animal welfare and steps taken to ameliorate suffering in all work involving non-human primates. The relevant guidelines followed and the committee that approved the study should be identified in the ethics statement. Please enter your ethics statement below	This research used publicly available imaging data not obtained at the authors' institutions. All subjects in the Open Access Series of Imaging Studies (OASIS; http://www.oasis-brains.org/) participated in accordance with guidelines of the Washington University Human Studies Committee. All subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI; http://adni.loni.usc.edu/) provided written informed consent and were recruited with respective institutional approval.

and place the same text at the beginning of the Methods section of your manuscript (with the subheading Ethics Statement). Enter "N/A" if you do not require an ethics statement.	

And so we came forth, and once again beheld the stars

Dante, Styron